ECONOMICS



THE MACMILLAN COMPANY
NEW YORK · BOSTON · CHICAGO · DALLAS
ATLANTA · SAN FRANCISCO
MACMILLAN AND CO, LIMITED
LONDON · BOMBAY · CALCUTTA · MADRAS
MELBOURNE
THE MACMILLAN COMPANY
OF CANADA, LIMITED
TORONTO

ECONOMICS

THIRD EDITION

Fred Rogers Fairchild
KNOX PROFESSOR OF ECONOMICS
EMERITUS

Edgar Stevenson Furniss
PELATIAH PERIT PROFESSOR OF POLITICAL
AND SOCIAL SCIENCE

Norman Sydney Buck
PROFESSOR OF ECONOMICS

YALE UNIVERSITY

New York
THE MACMILLAN COMPANY
1948

Third Edition Copyright, 1948, by THE MACMILLAN COMPANY

All rights reserved—no part of this book may be reproduced in any form without permission in writing from the publisher, except by a reviewer who wishes to quote brief passages in connection with a review written for inclusion in magazine or newspaper

Printed in the United States of America

One volume edition originally copyrighted and published, 1937; second edition, 1940, by The Macmillan Company

Preface

Many years' experience in teaching economics to beginners at Yale University has taught us the importance of the textbook and has led us to certain definite convictions as to the type of text best calculated to serve the purpose of helping college students to a knowledge of the fundamental facts and principles of the economic world in which they live. That is the sole purpose of this book. It is not written for the edification of the mature student of economics. It is strictly a book for beginners and the teachers of beginners.

Adherence to this purpose has indicated certain rules for our guidance and imposed certain restrictions upon us. As a rule the treatment has been confined to topics upon which the science of economics may fairly be said to have reached definite conclusions. Occasionally, when this rule could not be followed, we have stated —somewhat dogmatically perhaps—what we consider sound conclusions as to subjects upon which there may be real difference of opinion among economists of authority. In a few cases we have undertaken an impartial summary of considerations upon the two sides of some unsettled current issue, but as a rule we do not believe that the beginning student profits greatly from that type of discussion which merely presents arguments on both sides of a controversial question and leaves him to draw his own conclusions. Such matters belong in the more advanced courses. There is plenty of material to occupy fully the elementary course in economics without penetrating far into those frontier regions where economic science has not yet reached fairly definite conclusions.

This limitation of the field makes easier the strictly scientific treatment to which we have sought to adhere. Our purpose is to present the facts and the principles of economics—to disclose the truth. With the ethical or moral problems which arise upon this economic foundation we are not concerned, except only as prevailing judgments upon such matters may themselves become the

subject matter of scientific economic inquiry. Our purpose is to aid the student to discover how things are, not how they ought to be. We have no case to establish, and we offer no propaganda.

Addressed as it is to beginning students, this is essentially an elementary book. We have sought to spare the student the confusion that comes from taking for granted knowledge which he cannot fairly be assumed to possess, and the treatment is correspondingly elementary. We have sought, by giving care to such things as definition of technical terms, sequence of topics, and simplicity of style, to enable the beginning student to follow the discussion without undue difficulty. It is not to be denied that certain parts of the subject of economics are difficult. We have made no attempt to escape such difficulties, either by avoiding the hard topic or glossing it over with a superficial discussion. Rather have we relied upon simple and elementary treatment to remove unnecessary causes of confusion, in the belief that so presented these topics will not prove beyond the capacity of the student.

The plan of the book has dictated the continuous combination of theoretical analysis with historical narrative and discussion of practical problems. It is our conviction that this is a sound pedagogical rule for an elementary treatise. Theory must be illustrated and justified by showing its relation to practical affairs; facts must be made significant and interesting by showing their conformity to general principles. "Theory" and "practice" should, we believe, go always hand in hand.

It is today quite generally recognized that the average college student does not possess that knowledge of the economic environment which is necessary as a foundation for the study of economic principles and problems. A section (Chapters I–V) devoted to the study of the modern economic organization and its development, stressing the coöperative nature of modern society and the importance of price as a motivating force, is therefore made an integral part of this book.

The exercises which will be found at various points do not conform to any rigid rule. They have been introduced only at those points where it was felt that they would contribute substantial aid to the student in his effort to grasp the principles expounded in the preceding text or would give useful training in economic reasoning. These problems are obviously illustrative rather than

comprehensive, and the teacher can readily modify or expand them or ask the student himself to construct similar ones.

At the end of the book will be found a classified list of books which are suggested for the student's further reading. This list is of course not comprehensive, and it does not embrace works of a highly advanced or specialized or technical nature. It is intended merely to present for the choice of the student certain books which he will find interesting and profitable reading in connection with his first study of economics.

These in brief are the more important principles which have guided us in harmony with our aim to give to those students who may not pursue the study of economics further a broad knowledge and understanding of the economic world of today and to others a firm foundation for their further study of economics and allied subjects. It is in such matters as are here noted that our contribution, if any, is to be found.

While we have had in mind predominantly the needs of college students, we are nevertheless bold enough to believe that the requirements thus imposed upon us will have made this book useful also to those more mature readers who, for any reason whatever, may have need of a short elementary exposition of the fundamental facts and principles of economics in the light of present-day knowledge.

The present book is a shorter edition of our two-volume *Elementary Economics*. It has been written in response to the demand for an elementary textbook in economics similar in character to the longer work but of length suitable for use in half-year courses or in those full-year courses which use a general foundation text followed by selected special books. As regards purpose, balance of topics, plan, and style, the two books are generally alike.

During the past twenty or thirty years there has been a growing tendency, springing perhaps inevitably from the series of extraordinary events that crowded this period, to view with skepticism the principles of economics that found pretty general acceptance in the past, as well as a readiness on the part of many to throw these principles into the discard in favor of a "new economics."

The prosperous years from 1922 to 1929 were regarded by not a few persons as having ushered in a "new era", with poverty abolished and the human race lifted permanently to a level of prosperity

such as the mind of man had not before conceived. Economic science, casting doubts upon the validity of this roseate dream, was to be cast aside in favor of the "new economics" of perpetual prosperity. During the 1930's prolonged experience with adversity brought forth another "new economics", the economics of depression, which envisioned the necessity of continuous government spending on a vast scale if a huge body of unemployed were not to be placed permanently on government relief. More recently the world-wide maladjustments and dislocations in economic life created by a great war, together with the widespread poverty and insecurity that have followed in its wake, have engendered in many minds distrust of the capacity of a system of private enterprise to bring order out of chaos and to guarantee security to the masses, and the economics of authoritarianism or socialism has gained ascendency in many parts of the world where once economic liberalism and political democracy flourished.

It is our belief that a less superficial view discloses the prevailing body of economic doctrines as competent not only to explain the economic events of the recent past—whether of prosperity or adversity or postwar dislocation—but also to serve as a guide out of the economic confusion that grips the world today. We further believe that the basic principles of economics demonstrate the validity of the system of free enterprise, limited only by such government controls as are required to preserve it and make it work. It is in this spirit that the present edition has been prepared.

The following are some of the more important changes in this third edition. The chapters on price have been thoroughly reworked and rearranged, in some part to incorporate advances in theoretical knowledge, but in large measure to achieve greater simplicity and clarity of exposition. The chapters on banking have been recast in order to place greater emphasis on the way the banking system operates today. The chapters on government finance and taxation have been largely rewritten to deal more searchingly with prevailing ideas and policies bearing on the distribution of the tax burden and to recognize the increasing importance given to this phase of economic life by the depression and the second World War. Although the substance of the chapters dealing with the distribution of income have not been materially changed, the method of presentation has been considerably altered in an effort

to achieve greater unity in the laws of distribution. An entirely new chapter, *The Problems of American Agriculture,* has been added, and four others, though replacing some discussion of the corresponding subjects in the revised edition of 1940, are practically new: these are the chapters on *Principles of Taxation, Interregional Trade and Investment, Contemporary International Monetary Problems,* and *Transportation.* Finally, the entire book has been brought up to date with respect to statistics and all statements of fact.

In the prefaces to previous editions we have recorded the names of colleagues both at Yale University and at other institutions to whom we are indebted for assistance generously given in the preparation of various portions of this book, and we would once again acknowledge our gratitude to them, for their influence still lives in these pages. At this point however we shall not record again the names and specific contributions of those who have assisted us in past years, but rather refer the interested reader to the preface to the previous edition.

In the preparation of this edition we have called on many authorities in varied fields to help us, and it is a pleasure to express our gratitude to the following who have collaborated with us in an attempt to make this a more useful introduction to the study of economics. Professor Lloyd G. Reynolds, of Yale University, read and criticized the chapters on price, made special contribution to the treatment of price under imperfect competition, and rendered valuable assistance in the revision of the labor chapters.

Professor Arthur D. Gayer, of Queens College, contributed numerous useful suggestions for the revision of the chapters on money and banking.

Professor Michael A. Heilperin, formerly Associate Professor of Economics at Hamilton College, performed the major part of the task of revising the chapters on interregional trade and contemporary international monetary problems.

Professor R. G. Bressler, Jr., of the University of Connecticut, and the late Dr. Alan MacLeod, of the New England Research Council on Marketing and Food Supply, provided the substance of the chapter on American agriculture.

Professor Kent T. Healy, of Yale University, contributed the material for the revised chapter on transportation, and Professor

Joe S. Bain, of the University of California, for the chapter on government regulation of industry and trade.

The authors are particularly pleased to acknowledge their indebtedness to Professor Benjamin P. Whitaker, of Union College, for invaluable assistance in preparing copy for this revised and enlarged edition. Professor Whitaker assumed the task of incorporating into this revision the major changes which are to be found in the new fifth edition of our two-volume *Elementary Economics*. Certain sections of the book, designed to take account of the most recent statistical data and the latest developments in governmental policy, were prepared by him. Without his collaboration the materials for the book could not have been made ready for publication at this time.

Professor Whitaker's effort, moreover, was not one of editorship only, for he made many contributions to the thought as well as the form of this book.

We record also our indebtedness to Mr. Alfred E. Bobst, Jr., who ably served as Professor Whitaker's assistant.

It is to be understood that, in spite of the valuable contributions of these other economists, the authors assume full responsibility for everything in this book.

Fred Rogers Fairchild
Edgar Stevenson Furniss
Norman Sydney Buck

New Haven, Connecticut

Table of Contents

Chapter I. Foundations of the Science of Economics

1. Human wants: (1) Number and variety. (2) Food, clothing, and shelter. (3) Other wants. (4) Intermediate goods. (5) Man's wants unlimited
2. Limited means
3. Economics: (1) Definition. (2) Economics a science
4. Some fundamental concepts
 (1) Utility
 (2) Wealth: a. Definition. b. Are human beings wealth? c. How to identify wealth
 (3) Income
 (4) Costs
 (5) Net income
 (6) Wealth and income. Fund and flow
 (7) Property: a. Definition. b. Division of property. c. Property in free persons
 (8) Wealth and property
 (9) Documents in evidence of property
 (10) Value: a. A common denominator. b. Transfer and exchange. c. Barter and money exchange. d. Definition of value. e. Definition of price. f. Value and price. g. The unit of value
5. Exercises

Chapter II. Production 18

1. Production: (1) Creation of utility in wealth. (2) Direct personal services. (3) Definition of production
2. The factors of production
 (1) General
 (2) Land
 (3) Labor
 (4) Capital: a. Description. b. Capital in production. c. The accumulation of capital. Saving. d. The form of capital instruments. Invention
 (5) The entrepreneur
 (6) No corresponding class division
3. Division of labor: (1) Coöperation. (2) Division of labor. (3) Economies of division of labor. (4) Effects on labor. (5) Division of labor and the extent of the market. (6) Territorial division of labor. (7) Division of labor a form of coöperation
4. Exercises

Chapter III. The Business Firm 35

1. Function of the business firm
2. The individual proprietorship
3. The partnership
4. The corporation
 (1) Modern growth
 (2) Nature of the corporation
 (3) Capital stock
 (4) Bonds and notes
 (5) Stocks and bonds compared
 (6) The property account of the corporation: a. Assets and liabilities. b. Losses and insolvency. c. The value of stock. d. Interest payments and dividends. e. Stock dividends. f. Watered stock
 (7) The corporation's income account
5. Accounts of individual proprietorship and partnership
6. Exercises

Chapter IV. Large-scale Production and Combination 56

1. Rise of the standard of living
2. Growth of capital
3. Large-scale production
 (1) Importance of large plants
 (2) Conditions necessary for large-scale production
 (3) Economies of large-scale production: a. Management. b. Power and machinery. c. Technical experiment and research. d. By-products
 (4) The place of small-scale production
4. Importance of the large corporation
5. Combination: (1) Definition. (2) The formation of combinations. (3) Motives leading to the formation of combinations
6. Horizontal combinations: (1) Definition. (2) Economies in operation. (3) Economies in selling. (4) Economies in buying. (5) Saving in freight costs
7. Vertical combinations: (1) Definition. (2) Technical advantages. (3) The United States Steel Corporation
8. Conclusions

Chapter V. Economic Freedom and the Control of Production 75

1. The control of production
 (1) Conscious and unconscious control
 (2) Mercantilism
 (3) *Laissez faire*: a. Reaction from government control. b. Personal liberty. c. Private property. d. Free enterprise
 (4) Reaction from *laissez faire*
 a. General
 b. Motives: (a) To make freedom work. (b) To displace free enterprise
2. The present area of freedom
3. Viewpoint of this book
4. The free enterprise system: (1) The motive to productive effort. (2) Control of production by price. (3) Guiding production to meet con-

sumers' wants. (4) Allocating the factors of production among competing uses. (5) Rationing supplies of scarce goods. (6) Competitive bargaining

Chapter VI. Price and the Market 88

1. The problem of price: (1) Price in the free economy. (2) A pecuniary system. (3) Prices fixed by statute or custom. (4) Are sellers or buyers free to fix prices? (5) Is price a mystery?
2. The market: (1) Definition. (2) Free market
3. Why do people buy and sell?
4. Common and conflicting interests of buyers and sellers
5. Higgling exchange
6. The common auction
7. The buyer's auction: (1) Open bidding. (2) Sealed bids. (3) Price at the buyer's auction
8. Organized exchange
9. Price quoted by the seller
10. Price quoted by the buyer
11. Exercises

Chapter VII. Price Determination under Pure Competition 100

1. Competition
2. Pure competition
3. Where pure competition prevails
4. Imperfect competition
5. Determination of price on the organized exchange: (1) Price and demand. (2) Graphical representation, a digression. (3) Price and supply. (4) Demand and supply together. (5) Equilibrium of price and quantity. Clearing the market
6. The laws of price equilibrium
7. Alternative concepts of demand and supply: (1) Two possible meanings of demand. (2) Choice of a definition of demand. (3) Two possible meanings of supply. Choice of a definition. (4) The importance of consistency
8. The causes of price change: (1) In general. (2) Demand and supply the sole determinants. (3) The retail egg market. (4) The basic law of price change
9. Exercises

Chapter VIII. Demand 125

1. Introduction
2. Consumers' choices
3. Marginal utility: (1) An example. (2) Utility. (3) Marginal utility. (4) The problem of measurement. (5) Marginal utility curve. (6) The law of marginal utility. (7) Marginal utility of durable goods
4. Individual demand: (1) Marginal utility and individual demand. (2) Demand and desire. (3) Significance of this analysis. (4) How both parties to a trade can gain. (5) Real people. (6) Is all buying irrational?
5. Total demand: (1) Individual demand and total demand. (2) Definition of demand. (3) Time and place of demand schedule. (4) Actual demand schedules

6. The law of demand: (1) General statement. (2) Special cases. (3) Balancing marginal utilities. (4) Some apparent price anomalies resolved. (5) Abundance and scarcity
7. Elasticity of demand
 (1) General statement
 (2) Graphical illustration
 (3) Elasticity of demand and the amount of money spent
 (4) Practical examples. a. Inelastic demand. b. Elastic demand. Substitution
8. Exercises

Chapter IX. The Sellers and their Costs 154

1. The different viewpoints of buyers and sellers
2. Production to order or for a general market
3. Immediate and remote sales problems
4. Supply governed by cost
5. Future cost controls
6. Analysis of cost
 (1) Fixed and variable costs
 (2) Economy of fully utilized plant
 (3) Expansion of output and unit cost
 (4) The limit to large-scale enterprise
 (5) Marginal cost: a. Definition. b. Marginal cost and average cost
 (6) Relation of costs to output
7. Exercises

Chapter X. Supply 167

1. Marginal revenue
2. Cost of production and seller's policy: (1) Price above average cost. (2) Price below average cost. (3) Short run and long run. (4) Conclusion
3. Individual supply
4. Total supply
5. Supply and stock
6. Relation of quantity to price
 (1) Immediate sales problem
 (2) Remote sales problem
 a. Short run
 b. Long run. (a) Increasing costs. (b) Constant cost
7. The law of supply
8. Long and short run again
9. Elasticity of supply: (1) General statement. (2) Influences affecting elasticity. (3) The limits of elasticity
10. Decreasing cost industries
11. Effect of invention
12. Exercises

Chapter XI. Seller's Policy under Imperfect Competition 186

1. The conditions of pure competition. Resumé
2. The conditions of imperfect competition: (1) Ignorance, inertia, and agreement. (2) Individuality of products. (3) Small number of sellers. (4) Large offering of one seller

3. Imperfect competition defined
4. Oligopoly
5. A classification of market situations
6. The seller's problem under imperfect competition
7. Marginal revenue under imperfect competition
8. Marginal revenue and marginal cost: (1) In general. (2) Maximum profit. (3) Minimum loss. (4) Conclusions
9. The seller's liberty under imperfect competition: (1) Pure and imperfect competition compared. (2) Limitations upon the seller's liberty. (3) Liberty and discretion in big business
10. Quality and price in competition
11. The seller needs the facts of demand
12. Ascertaining the facts
13. Sales promotion
14. Advertising
15. Exercises

Chapter XII. Price under Imperfect Competition. Monopoly 211

1. Pure and imperfect competition
2. Principal bases of imperfect competition
3. Many sellers and a differentiated product: (1) What do we mean by "many"? (2) Effect of buyers' preferences. (3) Short-run problem. (4) Long-run problem
4. Imperfect competition with few sellers
 (1) The meaning of "few"
 (2) Why are sellers so few?
 (3) Short-run problem
 a. The demand curve
 b. Limits of price and output
 c. The seller's interest against price cutting
 d. Safeguards against price cutting: (a) Price leadership. (b) Agreements by sellers
 e. Economic obstacles to price control by sellers: (a) Unused capacity. (b) Marginal producers. (c) Conflicting interest of sellers.
 f. Conclusion
 (4) Long-run problem
5. Effect of fewness of buyers
6. Conclusions
7. Monopoly: (1) Definition and examples. (2) Monopoly price. (3) Monopoly and imperfect competition. (4) Monopoly and the public interest
8. Exercises

Chapter XIII. Interrelationships of Prices. Price Theory in Practice 231

1. Introduction
2. The present value of future income: (1) Capital and income. (2) Property and capital. (3) Present worth. (4) Present goods more desirable than future goods. (5) Interest. (6) The link between future and present values. (7) The discount process. (8) Capital and income. (9)

Discount and price laws. (10) Discount and the cost of production
3. Prices of the same good in different markets
4. Prices of competing goods
5. Prices of complementary goods
6. Prices of competing cost goods
7. Prices of joint cost goods
8. Prices of "tandem goods"
9. Price theory in practice: (1) Practical aspects of the theory of value. (2) The overworked term "value"
10. Exercises

Chapter XIV. Principles of Money 252

1. Nature of money
 (1) Some historical examples
 (2) The common characteristic
 (3) Definition of money
 (4) Functions of money
 (5) Government control of money: a. In general. b. Legal tender
 (6) Money and currency
2. Varieties of money and monetary systems
3. Standard money systems
 (1) Principal kinds of money
 (2) The ideal money commodity
 (3) Types of standard money systems
 (4) The gold standard: a. Characteristics. b. The monetary unit. c. Coinage. d. The value relation between gold and money
 (5) The gold bullion standard
 (6) The gold exchange standard
 (7) The bimetallic standard: a. Description. b. Operation of bimetallism. c. Gresham's law. d. Adjustment of discrepancy between ratios. e. Continued interest in bimetallism
4. Irredeemable money systems: (1) Origin and essential features. (2) General abandonment of the gold standard
5. Recent American monetary experience
 (1) Abandonment of the gold standard
 (2) Devaluation of the dollar
 (3) Gold reserve act and proclamation of 1934: a. Nationalization of gold. b. The new monetary unit. c. Legal tender. d. "Gold profit" and stabilization fund
 (4) Silver policy
6. The present American monetary system
7. Exercises

Chapter XV. Principles of Banking 280

1. Nature of credit
 (1) Financing production
 (2) Definition of credit
 (3) The function of credit
 a. Credit in general
 b. Bank credit: (a) Pooling of savings. (b) Creation of credit
2. Types of banking institutions

3. Loans and discounts: (1) Credit creation through lending. (2) Interest and discount. (3) Proceeds. (4) Principal types of trade paper. (5) A bank's balance sheet. (6) Discounting a note. (7) Paying a loan. (8) A typical bank statement
4. Bank deposits
 (1) Definition
 (2) Discount and deposit
 (3) The check
 a. Definition
 b. Payment by check. Within the bank
 c. Collection between banks: (a) In general. (b) Clearing house. (c) Clearing for outside banks. (d) The volume of clearings. (e) Out-of-town checks
5. The portfolio
6. Increasing importance of investments
7. The reserve
 (1) Its function
 (2) Control of reserve ration: a. Profits *vs.* safety. b. Means of control. c. Buying and selling securities. d. Rediscounting and borrowing. e. Calling loans. f. Control by the rate of discount
8. The normal business of banking
9. Note issue: (1) Similarity to deposits. (2) Differences between notes and deposits. (3) Legal restrictions
10. The problem of elasticity: (1) Need of elastic currency. (2) Elasticity of bank credit
11. Banking influence on economic fluctuations
12. Exercises

Chapter XVI. The American Banking System 313

1. Introduction
2. Historical resume: (1) Colonial banking. (2) The first Bank of the United States. (3) The second Bank of the United States. (4) Decentralized local banking
3. The national banking system: (1) Note issue. (2) Government depositories. (3) Reserves
4. The Federal Reserve System
 (1) Organization: (a) Board of Governors. b. Federal reserve districts and banks. c. Federal Advisory Council. d. Member banks
 (2) Functions of the Federal Reserve. General
 (3) Rediscounting and discounting
 (4) Note issue
 a. Federal reserve notes: (a) General description. (b) Security and reserve. (c) Loss of original elasticity concept
 b. Federal reserve bank notes. End of the national bank notes
 (5) Reserves against deposits: a. Legal requirements. b. Centralized reserves c. Mobile reserves
 (6) Control of credit: Qualitative and quantitative control. b. Control through rediscount rates. c. Open-market operations. d. Regulation of reserve ratios of member banks
 (7) National clearing of checks
 (8) Government fiscal agency

Chapter XVII. Recent Developments in American Banking 332

1. A decentralized banking system: (1) The local banks in 1929. (2) Branch banking limited. (3) Weaknesses of decentralized banking
2. Situation in the decade prior to 1933: (1) Expansion of credit. (2) Financial crises in 1929 and 1931. (3) A decade of bank failures
3. Structural changes since 1929: (1) Introduction. (2) Branch banking. (3) Holding company affiliates. (4) Capital requirements. (5) Divorce of commercial and investment banking. (6) Centralized control of the banking system
4. Government banking: (1) Reconstruction Finance Corporation. (2). Agricultural credit agencies. (3) Housing credit agencies. (4) Criticism of government banking
5. Functional changes since 1929
 (1) Control of interest on deposits
 (2) Deposit insurance: a. Description of the federal system b. Appraisal of deposit insurance
 (3) Loans on real estate
 (4) Speculative use of credit
 (5) Government control of credit
6. The present banking structure

Chapter XVIII. Money and prices 353

1. The value of money
2. Value of money not constant
3. Measuring the value of money
4. Index numbers: (1) Price relatives. (2) The average, simple and weighted. (3) An index number of the aggregative type. (4) Practical index numbers
5. The problem of the price level
 (1) Volume of trade
 (2) Quantity of currency
 (3) Velocity of circulation
 (4) The quantity theory of money: a. Algebraic relations. b. Significance of the quantity theory. c. Practical limitations of the quantity theory.
6. Far-reaching effects of monetary fluctuations
7. Monetary reform: (1) Stabilizing the value of money. (2) "Managed money". (3) The "Stabilized Dollar". (4) The "commodity reserve" scheme. (5) One hundred per cent money. (6) Status of monetary reform
8. Exercises

Chapter XIX. Demand for Factors of Production 374

1. Functional distribution: (1) Meaning. (2) The productive factors and their incomes. (3) Distribution occurs in the process of production. (4) Rates of interest, wages, and economic rent are market prices
2. Business demand for factors of production: (1) Motive. (2) Diminishing marginal productivity. (3) Law of entrepreneur-demand for productive factors. (4) Total entrepreneur-demand for productive factors
3. Non-business demand for factors of production
 (1) General

(2) Demand for labor
(3) Demand for capital: a. Consumption loans. b. Time preference. c. Borrowing for consumption as determined by time preference. d. Borrowing for governments. e. Demand for loans
(4) Demand for land
4. Laws of demand for productive factors. Summary
5. Functional distribution through the business world
6. The nature of profits

Chapter XX. Economic Rent 393

1. Introduction
2. The productivity of land
3. Agricultural lands: (1) Diminishing returns and marginal cost. (2) Marginal cost equated to market price. (3) Economic rent is producer's surplus. (4) Explicit and implicit economic rent. (5) Lands of different quality. (6) Margin of cultivation. (7) Alternative uses of agrarian land
4. Economic rent of urban lands
5. The value of land
6. Relation of rent to prices
7. Quasi rent
8. Economic rent distinguished from other income
9. "Unearned increment"
10. Exercises

Chapter XXI. Interest 410

1. Introduction
2. The loan market and the rate of interest
3. The supply of loan funds: (1) Savings of individuals. (2) Corporate saving
4. Supply of loan funds and the rate of interest
5. The rate of interest and the quantity of money
6. Lack of uniformity in interest rates
7. True and nominal interest rates
8. Changes in the rate of interest
9. Interest as a private income
10. Regulation of interest
11. "Over-saving" or "under-investment"

Chapter XXII. Wages 425

1. Approach to the problem of wage rates
2. The demand for labor
3. The supply of labor
4. The general law of wages
5. Limitations
 (1) Measurement of marginal productivity
 (2) Interdependence of different types of labor
 (3) The limits of bargaining: a. Wages above marginal product. b. Wages below marginal product
6. Variations in wage rates: (1) Individual variations. (2) Geographical variations in wage rates
7. Self-employed workers

8. Non-business workers
9. Importance of discounted marginal productivity
10. Wages and the contrast of social classes
11. Labor's share of the social income. The law of proportionality
12. Exercises

Chapter XXIII. Profits 441

1. Nature of profits: (1) The accountant's profit statement. (2) Implicit rent, interest, and wages of management. (3) Definition of profits. (4) Profits a residual income
2. Business risk
3. The risk-bearing function
4. Actual profits and losses
5. The historical "rate" of profits
6. Profits tend to be short-lived
7. Profits not a price determinant
8. Criticism of the profit system
9. Exercises

Chapter XXIV. Government. Public Finance. Taxes 452

1. Government's place in the economy
2. American political organization
3. The functions of government
4. Government income and costs
5. Financing government industry
6. Financing the primary functions of government: (1) General rule. (2) Taxes. (3) Fees. (4) Special assessments
7. Taxation, general concepts: (1) Technical definitions. (2) Relation of tax rate to base. (3) Another concept of progression
8. Principal forms of taxation
 (1) The property tax
 a. Evolution
 b. The American property tax
 (a) General description
 (b) Position of the property tax
 (c) Assessment: (a') The problem. (b') Expert assessment. (c') The ordinary town and county assessment
 (d) Practical results
 (e) Confusing wealth and property
 (f) Taxation of corporations
 (g) Reform of the property tax
 (2) The income tax
 a. Historical background
 b. The United States income tax: (a) General description. (b) Individual income tax base. (c) Corporation dividends. (d) Income tax rates. (e) Elasticity of the income tax
 (3) Consumption taxes
 a. Excise taxes
 b. The customs tariff: (a) General. (b) Economic effects of the tariff. (c) The revenue tariff. (d) Fiscal weakness of the protective tariff
 (4) Death taxes: a. General description. b. Death tax rates

(5) Business taxes
(6) The poll tax
9. Shifting and incidence: (1) General principles. (2) Incidence of excise taxes. (3) The customs tariff. (4) Taxes on economic rent and land value. (5) Capitalization of taxes

Chapter XXV. Principles of Taxation 482

1. An attribute of sovereignty
2. Sovereignty and taxing power in the United States
3. The purpose of taxation
4. The *quid pro quo* of taxation. In general
5. Distribution of the burden of taxation
 (1) "How ought taxes to be distributed?"
 (2) Tax distribution formulas
 a. Benefit or cost
 b. Ability to pay
 (a) What does it mean?
 (b) Equal sacrifice: Basic assumptions. Testing the assumptions
 (c) The other sacrifice theories
 (d) Is sacrifice "gross" or "net"?
 (e) Conclusion
 c. Just taxation. The rule of equity
 (3) Progressive taxation: a. Introduction. b. Is progression required by justice? c. Does ability to pay require progression? d. The degree of progression. e. "Equalizing fortunes." f. Prevalence of progression. g. Revenue aspects
 (4) A neutral tax system: a. Viewpoint. b. Narrowing the field of taxation. c. Proportional taxation. d. Relation to ability to pay. e. Summary and conclusions
6. Collateral effects of taxation
 (1) Geographical shifting of industry
 (2) Location of population
 (3) Form of tax
 (4) Extravagant business expenditures
 (5) Charitable contributions
 (6) Effect on utilization of land
 (7) Effect on industry: a. "Incentive taxation." b. Reducing the chance of profit. c. Small family businesses. d. Moderate *vs.* heavy taxation. e. Burden of compliance. f. Stability of law and administration
7. Nonfiscal taxation: (1) Sumptuary taxes. (2) "Protection." (3) Law by indirection. (4) Controlling the business cycle. (5) Arbitrary extremes

Chapter XXVI. American Public Finance 508

1. The cost of government
 (1) Public expenditure and national income
 (2) The record of public expenditure: a. General summary. b. Federal government of the United States. c. American state and local governments. d. Distribution betwen federal, state and local
 (3) Functional analysis

2. The revenue system: (1) Federal government. (2) The states. (3) Local governments
3. Public borrowing: (1) Purposes. (2) The security back of public debt. (3) High standing of government obligations. (4) The national debt of the United States
4. The budget
5. The balance of revenues and expenditures

Chapter XXVII. Interregional Trade and Investment 531

1. The bases of interregional trade: (1) General characteristics. (2) Causes of interregional specialization
2. The principle of comparative advantage
 (1) Absolute advantage
 (2) Comparative advantage
 (3) The principle of comparative advantage
 (4) Applications of the principle of interregional trade: a. Imports and exports. b. Multilateral trade. c. International prices. d. Industrialization and interregional trade
3. The structure of world trade
4. Foreign trade of the United States: (1) Volume. (2) Commodity character of foreign trade. (3) The direction of foreign trade. (4) Imports. (5) Exports
5. World trade after the Second World War
6. Foreign investments: (1) Their nature. (2) Purposes of foreign borrowing. (3) Purposes of foreign lending. (4) The transfer of payments. (5) The international investment position of the United States. (6) The International Bank for Reconstruction and Development

Chapter XXVIII. Foreign Exchange and the International Balance of Payments 555

1. Initial assumptions
2. Financing exports
3. Financing imports
4. Other international transactions
5. The rôle of banks in exchange
6. The exchange market
7. The rate of exchange: (1) General principles. (2) The gold points
8. Indirect exchange
9. Limits to the flow of gold: (1) General. (2) Effects of gold movements on price levels. (3) Gold movements and credit transfers. (4) Reaction of exchange rates upon imports and exports
10. The balance of payments: (1) Its composition. (2) The international balance sheet of the United States
11. Fallacies of the "favorable balance of trade"
12. Exercises

Chapter XXIX. Contemporary International Monetary Problems 575

1. Breakdown of the international gold standard: (1) Pre-war equilibrium. (2) Post-war exchange conditions
2. Irredeemable money and the par of exchange: (1) What is a "natural"

rate of exchange? (2) The purchasing power par of exchange. (3) Purchasing power par and market rate of exchange in a period of stability. (4) Purchasing power par and market rates in a period of inflation. (5) Limits to the usefulness of the purchasing power par concept
3. The return to the gold standard: (1) In general. (2) Stabilization of the pound
4. The gold exchange standard
5. Stabilization of exchange rates: (1) Flexible exchanges. (2) Injurious effects of flexible exchange rates. (3) Exchange control. (4) Official action to balance imports and exports with each country. (5) Exchange clearing systems
6. Measures to revive international monetary cooperation: (1) The "sterling area". (2) The exchange equalization system. (3) The Tripartite Agreement
7. The International Monetary Fund
(1) Objective
(2) The organization of the Fund
(3) Operations of the Fund
(4) Exchange stability: a. General provisions. b. Initial parities. c. The world price of gold. d. Exchange restrictions and freedom of international payments

Chapter XXX. Economic Nationalism and the Protective Tariff 595

1. International division of labor
2. National self-sufficiency
3. Effects of the protective tariff
4. Protection and national self-sufficiency
(1) The case for protection
a. The favorable balance of trade argument
b. The gold argument
c. The added industry argument
d. The wages argument: (a) Increased demand for labor. (b) Support of high wages. (c) Wages and labor cost
e. Alleged danger of withdrawing protection
f. Alleged saving in transportation costs
g, Alleged need to meet tariffs of other nations
(2) Considerations against protection: a. What is the territorial unit of protection? b. Benefits of foreign competition
(3) Conclusion
5. Temporary protection: (1) Infant industry protection. (2) Protection against "dumping"
6. State controlled trade
7. Protection on account of military necessity
8. Protection by bounties
9. Ship subsidies
10. International trade agreements: (1) Preferential arrangements. (2) Reciprocal trade agreements of the United States
11. War and restriction of foreign trade

Chapter XXXI. The Business Cycle 614

1. Business fluctuations. Character and types
2. Secular trend
3. Accidental irregularities
4. Seasonal cycles
5. The business cycle. Characteristic features
6. Varieties of business cycles
7. Analysis of the business cycle: (1) General. (2) Revival. (3) Prosperity. (4) Recession. (5) Depression. (6) Forces initiating the business cycle
8. Theories of the business cycle
 (1) The depression of the thirties
 (2) Theory of the matured economy: a. Statement. b. Criticism
 (3) The free economy: a. Recent restrictions. b. Pump priming and deficit spending. c. Restoration of investment. d. Stabilizing the economy

Chapter XXXII. Unemployment 640

1. Meaning of unemployment
2. Extent of unemployment
3. Causes of unemployment: (1) Demand and supply of labor. (2) Technological unemployment. (3) Seasonal unemployment. (4) Cyclical unemployment
4. Attitude of the laborers
5. Remedies for unemployment
 (1) Adjustment of labor supply: a. Mobility of labor. b. Public employment offices. c. Flexibility of wage rates
 (2) Stabilizing the demand for labor: a. Voluntary action by industry. b. Public works
 (3) Unemployment insurance: a. General nature and origin. b. The Ghent system. c. Compulsory insurance in Great Britain. d. Unemployment insurance in the United States

Chapter XXXIII. Other Labor Problems 661

1. Hours of labor: (1) The shorter workday and work week. (2) Hours and productivity
2. Child labor: (1) Characteristics. (2) Extent of child labor. (3) Effects of child labor. (4) Legal regulation of the labor of children
3. Labor of women: (1) Extent and character in the United States. (2) Causes of the increase in women labor. (3) Regulation of the labor of adult women
4. Wages of women
5. Minimum wage laws: (1) General summary. (2) The United States Fair Labor Standards Act of 1938. (3) Appraisal of minimum wage laws
6. Security for the aged
 (1) General
 (2) Old age pensions
 (3) Old age insurance: a. General. b. United States social security act
7. Industrial accidents
 (1) The problem
 (2) Distribution of the loss of work accidents: a. Individual loss. b. Common law remedy. c. Workmen's compensation laws

Chapter XXXIV. The American Organized Labor Movement 682

1. Structure of the labor movement
 (1) Conditioning factors
 (2) Types of American unions: a. Definition of labor union. b. The trade union. c. The industrial union. d. Conflict of rival types of unions
 (3) Strength of unionism in America
 (4) Legal status of unions: a. General situation. b. National Labor Relations Act. c. Labor Management Relations Act
 (5) Company unions
2. The program of American labor unionism
 (1) Political activity
 (2) Social features and benefit systems of unions
 (3) Collective bargaining: a. Purposes. b. Benefits of collective bargaining to employers. c. Conflict over methods in collective bargaining. d. Regulation of collective bargaining by the United States Government
 (4) The closed shop
 (5) Restrictive practices: a. Membership. b. Restriction of output
3. The strike: (1) Definition. (2) Statistics of strikes in the United States. (3) Types of strikes. (4) Violence in strikes. (5) Legality of the strike
4. Public intervention in strikes
 (1) The problem
 (2) Conciliation: a. Scope. b. Government encouragement of conciliation
 (3) Mediation: a. General nature. b. Official mediation in the United States
 (4) Voluntary arbitration: a. General nature. b. Official arbitration in the United States. c. Effectiveness of voluntary arbitration
 (5) Transition stages
 (6) Compulsory investigation
 (7) Compulsory arbitration: a. The affirmative argument. b. Experience in New Zealand. c. Compulsory arbitration in Kansas. d. The United States War Labor Board. e. Appraisal of compulsory arbitration

Chapter XXXV. Problems of American Agriculture 719

1. Economic peculiarities of agriculture
2. The American farm problem
 (1) Period of the First World War
 (2) During the 1920's
 (3) During the 1930's: a. Farm prices and income. b. Farm debt
3. Federal Farm Board
4. Farm programs of the Roosevelt Administration
 (1) General objectives
 (2) The Agricultural Adjustment Act of 1933: a. Objectives. b. "Parity prices." c. Production control. d. Processing taxes. e. Wheat programs. f. Other programs. g. AAA program and agricultural recovery. h. AAA program and business recovery
 (3) Soil Conservation and Domestic Allotment Act
 (4) Agricultural Adjustment Act of 1938
 a. Reasons for enactment

b. Principal provisions of the act of 1938:
(a) Soil conservation. (b) Commodity loans. (c) Parity payments. (d) Marketing quotas. (e) Crop insurance
c. General objectives of the act of 1938:
(a) Financing. (b) Income parity *vs.* price parity. (c) The "Ever-Normal Granary." (d) Crop Insurance. (e) Disposal of surplus commodities.
(5). Other agricultural programs
5. Farm program of the Second World War: (1) Production and prices. (2) Price "floors" and "ceilings." (3) Agricultural production during the war
6. Appraisal of farm policy: (1) Guiding principles. (2) Parity prices and incomes and (3) Conclusions

Chapter XXXVI. Transportation 744

1. Introduction. Economics of transportation
2. Topographical factors
(1) General
(2) Topography and the development of the American transportation system: a. Early development. b. Topography and the railroads. c. Topography and the modern highway system
3. Transportation costs
(1) General
(2) Fixed costs
(3) Variable costs: a. General. b. Utilization of vehicle capacity and variable costs
4. Competition between carriers
(1) Development of competition
(2) Restriction of the number of competing carriers
(3) Competition in rates: a. Conditioning factors. b. Inequalities due to competition and c. Competitive equalization of rates. d. Geography and rate competition
(4) Conclusions
5. Specific-rate determination
(1) Bargaining of shippers
(2) "Class rates"
(3) "Commodity rates"
(4) Rate discrimination
(5) Rate determination: a. Initiation of a rate change. b. Railroad rate associations
(6) Government regulation of particular rates
6. General level of rates
(1) Government regulation: a. Before 1920. b. The 1920 rule. c. The 1933 rule
(2) Historical review of general rate levels
7. The various forms of transportation at present

Chapter XXXVII. Government Regulation of Industry and Trade 769

1. Introduction
2. Types of government regulation

3. Antitrust laws: (1) The problem. (2) Provisions of the antitrust laws. (3) Intrinsic limitations of the antitrust laws. (4) Judicial interpretation of the antitrust laws and administrative policy. (5) Federal Trade Commission activities. (6) Collusive activity by export associations
4. The plane of competition: (1) The problem. (2) Pure food and drug acts. (3) "Unfair methods" of competition
5. Restraint of competition
 (1) Origins
 (2) National Industrial Recovery Act
 (3) Competition in the soft coal industry
 (4) The Guffey-Vinson Act to regulate the coal industry
 (5) Competition in the distributive trades: a. Rise of mass distribution. b. Chain-store taxation. c. The Robinson-Patman Act. d. Miller-Tydings Act. e. State "fair-trade" laws
 e. State "fair-trade" laws
6. Appraisal of governmental regulation
7. Government industry
 (1) Unquestioned types
 (2) Issues involved: a. Adequacy of government regulation. b. Savings on interest and profits. c. Corporation abuses
 (3) Relative efficiency: a. Financial results. b. Political interference. c. The motive in business. d. Conclusion

Chapter XXXVIII. Contrasting Economic Systems 796

1. Introduction. Present status of the free economy
2. Socialism
 (1) Theoretical basis
 (2) Socialist indictment of capitalism
 (3) Socialist analysis of capitalism
 a. Karl Marx
 b. The class struggle
 c. Trends in capitalist development: (a) Tendency toward concentration. (b) Increasing number and misery of laboring class. (c) The rôle of crises
 d. General appraisal of Marxian doctrine
 (4) The socialist program
 (5) The socialist campaign: a. Tactics. b. Political action
 (6) Appraisal of socialism
3. Communism
 (1) Its rise in Russia
 (2) The original communist program
 (3) Communist organization: a. Within Russia. b. International organization
 (4) Policies in Soviet Union: a. Recent history. b. Planned economy. c. Soviet economic organization. d. Growth of Soviet industry. e. Labor and wages in Soviet Russia. f. Soviet finance
4. Fascism
 (1) Present status
 (2) Comparison with communism: a. Similarities. b. Differences between fascism and communism

5. General appraisal of radical movements: (1) Authority versus liberty. (2) Planned economy and capital formation. (3) Planned economy and efficiency

SUGGESTIONS FOR FURTHER READINGS 833

INDEX 843

ECONOMICS

CHAPTER I

Foundations of the Science of Economics

Human wants: Number and variety. Those activities of mankind to which we ascribe the adjective "economic" and whose investigation calls for a *science of economics* are the consequence of two fundamental facts of nature; the one relates to the character of man himself; the other has to do with the natural environment in which man lives.

Organically man is an animal, one of the innumerable species of living things that populate the earth. Countless ages of evolution have brought man and all these other forms of life to their present state as separate species and their present equilibrium among themselves as sharers of the earth. It is especially to be noted that the process of evolution, while throwing man into the common struggle for existence, has developed in him certain characteristics which set him in a class apart from all other living things. Such peculiarity appears as soon as we give attention to the things that are desired by men.

In the earliest and simplest stage of evolutionary development, it is true, we seem to find nothing to distinguish man from the lower animals. In common with all forms of animal life, man has certain organic needs which must be met or he will die. He must have food and water, and air to breathe. His bodily temperature must be constantly maintained within very narrow limits, and he must, therefore, be protected from extremes of cold and heat by appropriate clothing and shelter. To protect himself against the attack of natural foes, man must have the means of defense or escape. These are fundamental organic needs, and the simplest means appropriate to meet them are called "necessities," employing the narrowest sense of this somewhat overworked term.

But, beyond this simplest stage, we recognize immediately a most significant difference between man and the lower animals. The latter are satisfied when the necessities of life are met. With man the necessities are only the beginning. No sooner are they met than

he reaches out for the satisfaction of other wants, for the "comforts and luxuries." Even the means whereby modern men satisfy the fundamental organic needs have passed far beyond the realm of bare necessities. This attribute of man is so fundamental to every phase of human life, and in particular to man's economic activities, that we must take pains to grasp its full import even at the risk of what may at first sight appear as an elaboration of the obvious.

Food, clothing, and shelter. Food is necessary to prevent starvation, of course, but that is not what the average person is thinking about as he seats himself at the dinner table. He is thinking of the joy of eating. He studies the menu. He rejects oysters and decides on clams. He scans the soup list. He ponders on the choice of roasts: chicken, turkey, beef, lamb; and so on down through the vegetables, salad, dessert, fruit, and coffee. This ceremony is something more than the warding off of starvation. And the same is true, only in lesser degree, of the simplest home table. To the original hunger instinct has been added a host of tastes and desires, so that we demand of food a good deal more than that it keep us alive. The strong desires that have developed from the drink-seeking instinct are too well known to need elaboration. Drinking is certainly not confined to the supplying of moisture to keep the body alive.

The necessity of protection from heat and cold gives rise to the desire for clothing. But to the modern civilized man or woman clothing means a good deal more than this. In buying a suit far more attention is given to the pattern, the cut, the style, the fit of the garments than to their cold-resisting qualities. Hats, coats, dresses, shoes, stockings, collars, ties, and so on down the list—how much of all this could be eliminated if all we wanted were protection against the elements!

In most climates some sort of shelter is a necessity, but the dwellings of civilized men, from ancient days down to the present, have been much more than this. From the humble cottage to the royal palace, men have sought to make of their dwellings, not mere shelters from the weather, but things of beauty, of comfort, of ostentatious display, of luxurious enjoyment. One will generally have to search long in the modern village or city to find any habitation that does not represent a craving for something beyond mere shelter. Our houses, like our clothes, are responsive to vanity. They also

represent (often grotesquely enough, to be sure) our aesthetic desire for that which is beautiful.

Other wants. The foregoing is, at best, only a beginning in listing the wants of mankind. Consider, for example, some of the things that the need of defense requires: strongly built houses, locks on doors and windows, protective weapons, policemen, armies and navies and their equipment, drugs and medicines, hospitals, jails and insane asylums, insurance companies. The gregarious tendency of mankind (the urge to work, play, and be with others) reminds us of the concentration of people in cities, towns, and villages, thus giving rise to special wants in the field of housing, transportation, etc. Even when people seek relaxation from their ordinary activities and surroundings and take vacations "in the country," they do not generally seek solitude. They require seashore and mountain hotels, resorts, and camps, where they can relax in company with others. Man has a thirst for knowledge, in its elements perhaps instinctive, but carrying him far beyond the promptings of mere instinctive curiosity. He wants books, telescopes, laboratories, schools, and colleges. He wants to travel and to visit new and strange scenes and so demands railroads, motor cars, ships, airplanes, hotels, and guides. He has a desire for the beautiful and so wants music and works of art. He enjoys sport and recreation and therefore requires golf clubs, baseball outfits, swimming suits, and all the paraphernalia of modern sport.

Intermediate goods. We must also note here the fact that modern civilized man wants many things which are not capable of satisfying directly any desire. The housewife wants a sewing machine, though there is obviously no satisfaction directly obtainable from the possession or use of this article. But the sewing machine will serve as an instrument for making and repairing the family clothing, thus helping to satisfy a fundamental need. It is not wanted for itself, but as a means to the gratification of this other need. This is a characteristic of the whole group of things known as instruments, tools, and machines, which make up a surprisingly large part of the commodities which men want and for which they strive. They are called "intermediate goods" and will demand further study in the next chapter and elsewhere in our investigation.

Man's wants unlimited. Such an enumeration of man's wants might be carried to any length. Enough has been said, however, to bring out the important conclusion with which we are at present

concerned; namely, the infinite number and the bewildering variety of the things that modern civilized man wants. Satisfying the instinctive cravings, securing the things necessary to sustain life and perpetuate the race—this is only the barest beginning. We soon forget that sustaining life is the basis of any of our wants. The more our wants are satisfied, the wider is the horizon opening up new and tempting vistas of wants still to be catered to.

We are quite justified in our conclusion that man's wants, as regards their number and variety, are without limit. This is the first of the two great facts which mark out the field of economics.

Limited means. The second fact is that the means for the satisfaction of human wants are limited. Man can obtain very little without effort on his part. Nature may be a bountiful provider; she is in fact the source of everything man has or can hope to have. Some of the things he wants, such as air, sunlight, water, she provides gratuitously. But after noting these few exceptions we conclude that nature does not generally bestow her gifts with a free hand. For the infinite number of things that man wants, man must work. Even the lower animals have to exert themselves to provide for their needs, and if man had been content with the sort of living which the lower animals thus obtain he would still be a lower animal himself.

Man is always engaged in efforts to secure the things he needs for the satisfaction of his wants. His wants are unlimited; the things to satisfy them are strictly limited. There is never enough to go around. He struggles ceaselessly to make more plentiful the things he wants; he develops continually ways and means of increasing his satisfactions from what he has.

Here is the cause of the greater part of the activities of human beings through all history. Man has always been chiefly engaged in "working for his living." Here are also the roots of the great human institutions—law, property, the family, etc.—which have developed out of man's struggle for the things with which to satisfy his wants.

Economics: Definition. The unlimited wants of man and the limited means available for their satisfaction are thus the foundation stones upon which rests the structure of economic science. Starting with these two great premises, economics investigates the activities of man in his efforts to satisfy, so far as he may, his limitless wants with the limited resources granted him by nature.

Economics is the science of man's activities devoted to obtaining the material means for the satisfaction of his wants. Lest the reader miss the true inclusiveness of this definition, he is reminded of what will be made clear a little later; *i.e.,* that the "material means" which serve to satisfy human wants include human beings as well as external objects.

When we consider that an important part, not to say the greater part, of all the waking hours of most adult men and women is devoted in some form or other to the business of acquiring the means for satifying wants, and that even the children are much engaged either in this same endeavor or in acquiring the education and training necessary to equip them to begin it a few years hence, we can realize that, although strictly limited by the definition, the field of economics is a broad one. We might proceed to catalogue the many topics included and the various problems to be investigated. But a mere glance at the table of contents of this book will perhaps suffice to give the reader a general notion of the subject matter of economics.

Economics a science. To designate economics a science not only imposes definite limitations on the scope of its subject matter but also prescribes the kind or quality of analysis involved in economic investigations. Broadly, a science may be defined as a body of knowledge in some particular field developed and arranged in an orderly way to show the relations existing among its facts or parts. Science deals with facts and their relations. Its function is limited to the discovery and enunciation of the relations existing among facts.

It is not the function of science to tell how things ought to be, to lay down moral precepts, to advocate that which is good, or to condemn that which is evil. Science undertakes to tell how things are, not how they ought to be.

Nor is it the business of science to lay down rules for the accomplishing of desired ends. That is the function of an art. Thus the sciences of astronomy and mathematics provide certain facts and relations between them; these are employed by the art of navigation to accomplish its desired ends. The sciences of physics and biology are similarly availed of by the art of medicine.

Science is concerned only with discovering and stating the truth. It leaves to others the use of its results to influence public opinion or government action. Science neither condemns nor commends;

it explains. The science of astronomy furnishes the knowledge that enables the navigator to bring his ship into the desired port, but astronomy expresses no judgment as to whether the ship should have been brought to this port rather than somewhere else.

In thus sharply marking out the field of science we do not imply that moral or ethical considerations are to be ignored by the student of economics. Economics is a social science, and as such it grasps as its subject matter all attributes of man which affect his pursuit of the material means for the satisfaction of his wants. Among these attributes the moral and ethical notions of mankind are of importance and require full consideration. For example, the moral code of one community may dictate polygamy, while that of another insists upon monogamy, and these moral edicts are of great economic significance. The science of economics will tell us a great deal about the respective effects of these two systems of family organization, but we must never forget that economics cannot and does not attempt to tell whether monogamy or polygamy is right or wrong.[1]

It follows that, as a science, economics does not seek to persuade people to accept certain beliefs or to adopt certain codes of conduct. In a free country like America, there are innumerable "causes": monetary reform, free trade, protective tariff, collective bargaining, social security, free enterprise, economic planning, international cooperation, universal military training, and so on. Those who engage in promoting such causes are acting as advocates, not as scientists.

Some fundamental concepts: Utility. We need now to acquire clear notions of certain fundamental concepts which are inherent in the subject matter of economics and are an essential part of the equipment of the student.

Man desires those things that are able to satisfy his wants. Anything that satisfies a human want is useful; it has *utility*. *Utility is that quality of a thing which enables it to satisfy a human want.* This definition scarcely needs elaboration or explanation. We determine readily enough whether we want any particular thing; *i.e.*, whether it is "useful" or "useless."

As a science, economics gives somewhat broader meanings to "useful" and "utility" than are usually employed in common

[1] Cf. Walker, F. A., *Political Economy*, Third Edition. New York: Henry Holt and Company, Inc., 1888. Pp. 19-21.

speech. When the economist calls anything useful, he implies nothing as to its merits in an ethical sense. A thing is useful if it can satisfy a human want; that is all. He does not inquire whether the satisfaction of that particular want may be good or bad, wise or foolish. Opium, except when used as a medicine, may be physically injurious and morally bad, but this does not alter its classification as an object of utility so long as it does actually enable a human want to be satisfied. The use of chewing gum is probably foolish, but we cannot thereby deny its utility. The pictures which adorn the walls of many homes may be wholly bad from an artistic sense, but they are nevertheless economically useful. Of course this does not imply any lack of appreciation of the importance of aesthetic and ethical judgments. Such considerations are simply outside the field of economics.

Wealth: Definition. *Wealth consists of all useful, material things owned by human beings.* This definition needs some further examination. We have already settled the meaning of "useful." Useless things are not included in wealth, which thus embraces only those things which are capable of satisfying human wants. The reader should perhaps again be warned that the ethical character of human wants does not enter into the classification.

We limit wealth secondly to material things. This excludes such desirable things as honesty, good health, the speed of a race horse, the coldness of ice, the skill of an artist or a musician. This may possibly seem arbitrary and unfortunate at first. But we shall not have to look far to see the great advantage of this limitation and particularly the confusion which would follow if the concept of wealth were not thus limited. If the speed of a great race horse is wealth, what of the horse himself? Certainly he is wealth. But have we then two articles of wealth, the horse and his speed? And if so, should we not add a third, his strength; and a fourth, his beauty?

Most of the forms of "immaterial wealth" with which we should have to deal if our definition admitted them are simply qualities of material things which are themselves wealth. To count both the material thing and its qualities as wealth would be productive of nothing but confusion. Such double counting is really to confuse wealth and utility. Similar confusion would arise from including human qualities, such as health, strength, skill, honesty, in the category of wealth. We shall consider a little later the question whether human beings themselves are wealth.

Having excluded from our concept of wealth all useless and all immaterial things, we next limit it to those useful, material things which are owned by man. The first effect of this limitation is to exclude those useful things which are furnished freely and abundantly by nature. Thus air is material, and it is certainly one of the most useful things in the world. But on account of its abundance it is not owned. Why should one take the trouble to own it? We also exclude certain things which, from their nature or from the nature of the human institution of ownership, cannot be owned. We cannot own the sun or the moon; no person or group of persons can own the Atlantic Ocean or the Gulf Stream.

Most material things that satisfy human wants, however, are owned. Ownership may be vested in the individual person, as in the case of one's suit of clothes, the farmer's land, the merchant's store, etc. Or various groups of persons, such as partnerships, corporations, etc., may own land and buildings and virtually any kind of wealth which may be owned by individuals. Governments—national, state, and local—are large owners of wealth; they are simply another form of association of human beings.

Are human beings wealth? The requirement of ownership by human beings restricts the scientific concept of wealth in another important respect. Are human beings themselves wealth? As regards slaves, the question is clearly answered in the affirmative by reference to our definition. The slave is useful; he is a material thing; he is owned by his master.

But how about free men or women or children? Free persons are material things, of course, and they have qualities which enable them to satisfy wants of other persons. But they are not owned under the legal systems of most modern nations. It is not always easy to know just where to draw the line between free men and those who are owned. We might set up the fiction that free persons are owned by themselves or (in the case of minor children) by their parents. Free persons would then fall into our category of wealth, and there are some advantages to this classification. Logic and simplicity would perhaps be served by including all human beings in the category of wealth.

On the other hand, it is not customary to think of free persons as owned, even by themselves, nor do we naturally think of free persons as wealth. The weight of advantage is on the side of our definition as stated, according to which, while slaves are undoubt-

edly wealth, free persons are not. However, although free persons are not wealth, they do render services possessing utility similar to those rendered by articles of wealth, and we shall have to take account of these services in our study of economics.

How to identify wealth. To determine whether any given thing is wealth or not, we have only to ask three questions: Is it useful? Is it material? Is it owned? If the answer is yes to all these questions, the thing is wealth. A negative answer to any one of the questions excludes it.

Income. When an article of wealth satisfies a want of man we say that it furnishes *benefits* or renders *services.* By the benefits or services of wealth we mean *desirable events which it causes for human beings.* Keeping out the cold is a desirable event caused by a dwelling house; it is one element of the service of shelter, a benefit yielded by the house. The service of a piece of ice in the refrigerator is keeping things cold. A diamond ring renders service by satisfying the love of beauty or perhaps only the vanity of its wearer. The service of a slave is working in the field or as a domestic servant.

Thus every article of wealth renders its appropriate service to man, else it would not be useful and so would not be wealth. In exactly the same way services are rendered by free persons, as when an actor performs before an audience or a lawyer tries a case for his client. The word *income* is used to denote such services and is defined as follows: *Income consists of the benefits or services rendered by wealth or by free persons.*

Costs. Not all the events caused by wealth are desirable. Besides furnishing shelter, a dwelling house requires painting and repairs; it compels its owner to pay taxes and insurance premiums, and so on; finally it wears out or becomes obsolete and if not discarded must be torn down and rebuilt. The slave works in the field, but he requires food, clothing, and shelter, and the free farm laborer requires wages. *The undesirable events caused by wealth or by free persons are called disservices or costs.*

Net income. All wealth yields both income and costs. We cannot escape the undesirable events if we would have the desirable events; we put up with the costs in order to get the income of wealth. It follows that, in the long run, the income from any article of wealth must exceed its costs (how these are measured

will appear later), else no one would care to own that article, and it would cease to be wealth.

For example a storekeeper has a motor truck to deliver goods to his customers. The truck is an article of wealth; the income from it is the service of delivering goods or the receipt of money for delivering goods; its costs are the expenditures of money for the gasoline and oil it consumes, garage room, occasional washing, and repairs. These costs would not be borne if the service of delivering goods were not worth more than the costs. Eventually the truck begins to wear out; its speed diminishes and it cannot make some of the grades. It consumes more gasoline, and its loose piston rings and bearings are wasteful of oil. Repair bills become heavier and more frequent. Thus its income declines while its costs rise, and sooner or later the margin between them disappears. The machine is no longer useful, and it finds a resting place on the junk heap; it is no longer wealth.

The difference between the income and the costs of any article of wealth is its net income. Some net income must be present or the article will not be wanted. In general, then, utility means the power to yield a net income or an income in excess of the costs, and this idea of a net income is inherent in the term "useful" as employed in the definition of wealth.[1]

Wealth and income. Fund and flow. There is a fundamental difference between wealth and income, which manifests itself in the matter of measurement. When wealth is measured, it is the quantity in existence at a certain time that is determined. In measuring income, we ascertain the amount of service that is rendered during a certain period of time. We do not speak of the yield of an orchard or the service of a carpenter at 1:25 P.M. of a certain date. An orchard yields so many bushels of apples per year, a mason lays so many bricks per day or per hour, etc. Wealth is a *fund;* it is measured as of a certain *instant* in time. Income is a *flow;* it is measured as of a certain *period* of time.[2]

[1] The reader who has any familiarity with accounting will note that these definitions of income, cost, and net income differ somewhat from those used by the accountants. They are not really in conflict. The accountant's terminology represents a special technical application of the more general concepts with which we are here concerned.

[2] By the methods of the infinitesimal calculus, the period taken for the measurement of income may be reduced to zero, thus making possible the measure-

Property: Definition. Every article of wealth is owned by somebody. This relation between wealth and its owner is one of the fundamental concepts of economics. It may be called ownership, property right, or property. *Property* is the technical term generally used in economics. It is defined as *the right to income; that is, the right to the benefits or services of wealth or free persons.* This evidently is what ownership means. When a person owns an article, he has a right in it; namely, the right to have it and use it, all of which is included in the right to the services or income from that article. The owner is protected in his property right by the laws and customs of the community in which he lives; other persons are prevented from interfering with his enjoyment of the services of his wealth.

Since the term "income" is so broad as to include any use to which the owner could put his wealth, including even its destruction if he so desires, property might be defined simply as the right to wealth or free persons, without any change in meaning. The definition here chosen has the advantages: (1) of emphasizing the fundamental importance of income, (2) of somewhat readier adaptation to the innumerable cases of divided ownership, and (3) of avoiding any apparent conflict with the common usage of not regarding free persons as owned.

Division of property. A sole owner, such as the sole owner of a house and lot, an automobile, or a suit of clothes, has the exclusive right to all the income of the article of wealth which he owns. Ownership may, however, be divided, in which case each of the two or more owners has certain rights in the wealth; that is, the right to certain of its benefits or income. The tenant of a rented house has the right to its use only during a certain time, as limited by his lease contract with the landlord, who has retained the right to the services of the house at all other times; that is, before the beginning and after the termination of the lease. In a partnership the wealth held belongs jointly to the two or more partners. Each has the right to a certain part of the services or income of this wealth, the exact division being specified in the partnership agreement.

So in the corporate form of business organization, the wealth of

ment of income as a *rate* of flow at an instant of time. This concept is of great importance in certain advanced problems of economics, as it is in some of the natural sciences. It need not further concern us in this elementary book.

the corporation belongs in part to the stockholders, of whom there may be hundreds or thousands, each with certain rights in the income from the corporation's property. The bondholders of the corporation represent another group of part owners, with certain other rights in the corporation's wealth, all specified in the bonds. Thus each bondholder has the right to receive certain interest payments at certain specified dates and finally at a specified date to receive a certain larger sum equal usually to the principal amount stated in the bond. That part of the income from the corporation's property which remains after the rights of bondholders and possibly other classes of part owners, such as noteholders, etc., have been satisfied belongs to the stockholders and may be distributed to them in the form of dividends.

Property may thus be divided in a great variety of ways. Every person who has a property right in an article of wealth is an owner (sole or part) of that wealth, though in common speech the term "owner" is sometimes applied to only one of the parties. Thus the landlord and the tenant are really both owners of the rented house. Neither is the exclusive owner, for neither has the right to all the services of the house. The property is divided between them, though we are accustomed to apply the term "owner" to the landlord alone. In like manner the bondholders are fundamentally part owners, although not usually spoken of as such.

All the services of wealth must belong to somebody. When property is divided, the sum of all the property rights of the several part owners must equal the total property right to all the services of that piece of wealth. For every article of wealth there must be a property right or property rights, and back of every property right there must always be an article of wealth or a free person.

Property in free persons. The last words of the previous paragraph will recall that property may be a right to the services of free persons as well as of wealth. If we had chosen to include free persons in the definition of wealth, all that has been said of property in wealth would have applied equally to them. As it is we must recognize that free persons render services to others and that these others have rights to such services.

A baseball player signs a contract with a manager agreeing to play exclusively for him during the coming season. The manager has thereby a property right, the right to receive certain services from this player. To that extent the player's right to his own time

and services—actually his right to himself—is diminished. It is divided property again. Theatrical managers, moving picture producers, musical promoters, and others have property rights in actors and actresses, musicians, and other artists. Every employer has a property right to certain services of his employees. A promissory note gives the creditor a property right in the debtor; that is, the right to receive a certain sum of money. Lawyers have rights against their clients, and clients against their lawyers; parents have rights in their children, and children in their parents. Every business contract gives each party a property right against the other.

In fact the property in persons is almost always divided. Practically no one is completely his own master, and there are all degrees of property rights of others, from the slave, a true article of wealth because completely owned by his master, through serfs, peons, indentured servants, to the freest so-called "free man."

Wealth and property. In popular speech, the terms "property" and "wealth" are not carefully distinguished. It is quite usual, for example, to speak of a house and lot as "property," to call household furniture "personal property," and so on. For the purposes of scientific analysis, economics draws a sharp distinction between property and wealth. The latter is a physical, tangible thing; it is material. Property, on the other hand, is a relation between wealth (or free persons) and persons; it is immaterial.

This distinction is fundamental and has many important bearings. Thus wealth is the source of income, whereas property is the distributor of income. The wealth of a community, together with its productive free persons, determines how much and what kind of income the people may enjoy. The property rights of the several members of the community determine how these incomes shall be divided among them, how much of the total community income each one shall enjoy. Property rights may be changed without changing the total amount of wealth. The material welfare of the human race is a matter of the satisfaction of wants and is affected both by the total quantity of wealth and by the way the property rights to wealth are distributed among individuals, families, and nations.

Documents in evidence of property. Property rights are sometimes expressed in legal documents. The owner of a farm or a building lot has a *deed,* which is a piece of paper containing a legal statement of his property right to the benefits of a specified

parcel of land. When one purchases the right to sit in a certain seat of a theatre on a certain date, his property is evidenced by a ticket. On the other hand, no legal document testifies to one's property right in the clothes he wears or the food on his table. Whether there shall be a document or not is a matter of custom or law. The property is the original and fundamental thing; the document, when there is one, is secondary and incidental; in many cases no document is considered essential.

The reader must be on his guard against confusion between property rights and the documents in evidence thereof, particularly in those cases where common, and even scientific, usage has not found it necessary to employ different terms for these two things. For example, a bank note is, strictly speaking, a piece of paper testifying to a property right of the bearer against a bank. Yet it is common to speak of the bank note as though it were the right itself. When once he has been warned, the reader is not likely to have any difficulty with such usage, loose though it confessedly is.

Value: A common denominator. Each kind of wealth, income, or property has its appropriate unit of measure. But there is one unit in which all kinds of wealth, income, and property may be measured and which is by far the most important of all units. This is the unit of value or price.

Transfer and exchange. When Mr. Jones gives his dwelling house to his wife, there is a change of ownership. The house, which formerly belonged to Mr. Jones, now belongs to Mrs. Jones. A *change of ownership of wealth* is called in economic terminology a *transfer*. Transfer takes place whenever wealth is bought or sold, bequeathed, or given away.

Now suppose Mrs. Jones sells the house to the Brown Realty Co. for $20,000. There has been another transfer of the house. But this time there has been also a transfer of money, from the Brown Realty Co. to Mrs. Jones. These two transfers together make an *exchange,* which is defined in technical language as *a pair of voluntary transfers between two owners, when each transfer is made in consideration of the other.* A gift is not an exchange, since there is only one transfer. As is said, there is "no consideration."

Barter and money exchange. *Exchanges made without the use of money are called barter.* For example, the farmer brings ten dozen eggs to the village storekeeper and takes in return fifteen gallons of gasoline and five quarts of motor oil. This transaction

fits the definition of an exchange. "Swapping" horses or anything else is barter. At some time in the history of any people barter was probably the usual form of exchange, such as there was, and a certain amount of barter goes on among all peoples today.

But the civilized world long ago abandoned barter as the normal form of exchange. In practically all exchanges today one of the transfers is a transfer of money, and money exchange has thus taken the place of barter. The universality of exchange is one of the distinguishing characteristics of modern economic organization. In fact, because of the dominance of money exchanges, that organization is frequently called a money economy, and money, as we shall find in later chapters, is defined as *anything which is generally accepted in exchange for other things.*

Definition of value. From exchange thus come two of the most important concepts of economics and business, not to say of modern human life in general; namely, value and price. As the term is generally employed in economics, *the value of anything is the quantity of any other thing that would be given in exchange for the first thing.* If ten dozen eggs could be exchanged for two bushels of potatoes, the value of the eggs is two bushels of potatoes, and the value of the potatoes is ten dozen eggs. If a ton of stove coal is sold for twelve dollars, the value of the coal is twelve dollars, and the value of twelve dollars is one ton of stove coal. It is not customary, however, to speak of the value of money in this way. On the other hand, values of practically all other things are regularly expressed in terms of money.

Definition of price. Price is a more limited, and, therefore, a more specific, concept than value. First, it is always expressed in terms of money; and secondly, it always relates to one unit of the thing measured. As so defined, *the price of anything is the amount of money that would be given in exchange for one unit of it.* If a pound of cotton would sell for twenty cents, the price of cotton is twenty cents per pound. In stating the price of anything, the unit of measure used must always be stated or understood.

Value and price. The terms value and price are of course used more or less vaguely and inconsistently in popular speech, and there is some difference in usage among economists. The definitions here given are in good standing among economists, they agree as closely as is possible with popular usage, and they will prove useful for economic analysis.

As here defined, value and price are closely related to each other. They are distinguished from each other in that (1) value may be expressed in terms of any kind of wealth, property, or service, while the price of anything is expressed only in money, and (2) the term "value" is used for any quantity of a good, whereas price relates specifically to one unit. In practice, value as well as price is usually expressed in terms of money; in which case value is price multiplied by quantity, and the value of one unit is identical with the price. Thus if a farm sells for $25,000, we may say that the value of the farm is $25,000, or that its price is $25,000. We should not say that the price of fifteen tons of coal is $150, but should say that the value of fifteen tons of coal is $150, because the price is $10 a ton.

The unit of value. Value thus gives us a common unit in which to measure all kinds of wealth, income, and property. This is the unit of money value; *i.e.*, the value of the monetary unit, in America, the dollar. Since value is almost always expressed in terms of money, we may ordinarily call the dollar the unit of value, not bothering always to say the unit of "money value."

Thanks to this common unit, it is possible to total a list of various articles of wealth and property rights. Bushels of potatoes, yards of cloth, pounds of sugar, acres of land cannot be added together. But the values of all these things can be. It is possible therefore for the business man to make an "inventory," for the citizen to make up his "tax list," for anybody to make a list of his possessions and get the total. Above all, it is possible to keep accounts of wealth and property and income.

EXERCISES

1. Which of the following may properly be called wealth? Explain in each case.
 (*a*) A fountain pen. (*b*) An abandoned farm. (*c*) A factory building. (*d*) A state highway. (*e*) The Gulf of Mexico. (*f*) The salt water of the Gulf of Mexico used in the production of magnesium. (*g*) A farm whose ownership is under dispute. (*h*) Fish in the ocean. (*i*) A moving picture actor. (*j*) A college dormitory. (*k*) An automobile in the junk yard. (*l*) A coal mine.
2. Which of the following are wealth and which represent property rights? Explain in each case.
 (*a*) A share of stock. (*b*) A theater building. (*c*) A bank deposit. (*d*) A dollar bill. (*e*) A suit of clothes. (*f*) A life insurance policy. (*g*) A de-

livery truck. (*h*) A football ticket. (*i*) An apple orchard. (*j*) A mortgage on a residence. (*k*) A state park.

3. During the recent World War the United States Government borrowed billions of dollars from the people and from banks through the sale of bonds.

 (*a*) Does this debt affect the wealth of the United States? (*b*) During the years these bonds are outstanding, interest will be paid the bondholders out of funds raised by taxation. Does the annual payment of interest on a national debt of this kind affect the wealth of the country? (*c*) Would the wealth of the country be affected if the bonds were paid off? (*d*) Would the wealth of the country be affected if the government arbitrarily repudiated the debt?

 Note: This exercise refers to immediate effects, not to remote or indirect consequences.

4. In 1932–33 a number of states adopted laws providing for moratoria on farm mortgage indebtedness. The holders of mortgages were unable to collect interest or undertake foreclosure.

 (*a*) Did such laws affect the wealth of these states? (*b*) Did such laws affect property rights?

5. Describe the income created by each of the following.

 (*a*) A farm. (*b*) A musician. (*c*) A watch factory. (*d*) A doctor. (*e*) A baseball player. (*f*) A fishing vessel. (*g*) A housewife.

6. Which of the following can be described as costs? Explain in each case.

 (*a*) Wages of a janitor of an apartment building. (*b*) The damage caused by the smoke from soft coal used in heating the apartment building. (*c*) Real estate taxes. (*d*) Income taxes on the net income derived from operating the apartment building.

 What is meant by the expression, "to secure a bargain"? Reconcile the popular interpretation of this phrase with the discussion of value and price.

CHAPTER II

Production

Production: Creation of utility in wealth. Economics has been defined as the science of man's activities devoted to obtaining the material means for the satisfaction of his wants. That part of these activities most readily identified we refer to as production. But what specifically is production? Who are the producers of the world?

Common usage would lead us to say at once that the farmer, who raises wheat, is a producer; he obviously produces wealth, in that he creates something which did not previously exist. But a little reflection will show that he is a creator or producer only in the sense that he has been instrumental in effecting a change in the form or relationship of various physical elements already in existence so as to make them suitable for human consumption. The cotton manufacturer likewise adds nothing to the world's store of materials, but when he twists cotton fibre into thread and weaves thread into cloth he has added to the utility of the cotton in the eyes of the consumer. If we examine other forms of agriculture or manufacturing, we shall find that the essence of the particular operation is to bring about a change in the form of the good which will bring it one stage nearer the condition where it may satisfy a human want. Agriculture and manufacture, carpentry and blacksmithing, and innumerable other lines of activity, by thus changing the form of material goods, are active in the *creation of utility.*

But this is not the only way that utility is created. There may seem to be little similarity between farming and the operation of a railroad if one views only the technique of each business, but the purpose is essentially the same. Wheat on the farms of the West is quite as useless to the consumer of the East as if it had never been grown, and the service of the railroad in bringing it to the consumer is just as essential as that of the farmer in growing it.

The grocer who keeps flour in stock to meet his customers' demands is performing a similar service, for the consumer not only

wants a particular good at a particular place, but he wants it at a definite time. Today's flour is patently more useful in satisfying today's hunger than flour which will not be available until next week.

Likewise a productive service is rendered by those who facilitate the sale of goods. When A desires to sell wheat, he gives his broker a warehouse receipt for wheat stored in some elevator; the broker sells the wheat to B by delivering to him the warehouse receipt. A service has been rendered, for which the broker will collect commissions from A and B. The wheat is presumably one step nearer to the direct satisfaction of a human want. Utility has been created, by a change in ownership, even though the actual wheat has not been touched.

Thus are illustrated the four forms which the production of wealth may take—the creation of *form utility,* of *place utility,* of *time utility,* and of *ownership utility.* All effort directed toward the production of wealth may be classified under one or another of these four headings: agriculture and manufacturing result primarily in the creation of form utility; all the varied transportation services are engaged in creating place utility; merchants of all classes, bankers, brokers, and other middlemen are creating time utility or ownership utility. To the first one alone the name production is sometimes given, doubtless because of the obvious character of the change, but in view of the essential nature of the other three forms it is difficult to justify such usage.

Direct personal services. Although our primary interest in the study of production centres around the production of wealth, it should not be forgotten that man has wants which cannot be satisfied by wealth alone. A mechanical piano may satisfy the longing of one for music, but to another it is a sorry substitute for the performance of the pianist. If the labor of the pianist in making the roll for the mechanical piano is productive, certainly his performance on the concert platform must also be productive. We must therefore include in our definition of production the rendering of direct personal services by free persons.

Definition of production. Completely stated, *production consists of (a) the production of wealth, which is defined as the creation of utility in wealth, and (b) the rendering by free persons of direct personal services.*

This definition is intentionally broad, including any effort which

results in the satisfaction of a human want. In defining utility, it was pointed out that scientific definitions have nothing to do with questions of ethics or morals. Like utility, the word production carries with it no moral significance, although economists are not therefore unaware that effort spent in some directions may be less commendable than effort spent in others, and that the results of human exertion may be fatal to individuals or disastrous to society. The study of production is the study of human activity in satisfying wants, whether or not that activity can be approved on moral or ethical grounds.

The factors of production: General. In his struggle to guide and control natural forces to his own ends man has enlisted in his service two powerful partners, nature and capital. These three—man, nature, and capital—work together in production, with results which, when compared with man's first efforts, are almost unbelievable. These three partners are commonly called the *factors of production* under the titles, *land, labor,* and *capital,* the first two being figuratively used to stand for nature (not merely land in the narrow sense) and man.

Land. Land is essential in providing "standing room" for the population of the earth; it is the source of the food supply for the people and the raw materials for production and is the original source of all that man has. It is more than mere land, for it embraces the whole natural environment, including the oceans, lakes, and rivers, the topography of the land, the character of the soil, the amount and seasons of the rainfall, the range of temperature at different seasons, the mineral deposits, the water power, and a host of other features, all of which are important and may have a determining effect on the character of occupation and the type of industry in a particular region.

Labor. Land without labor is useless for the purposes of man. Even in the spots most favored by nature some effort is required to secure the goods desired, and on most of the earth's surface much exertion is necessary to make a living. For the majority of men labor is painful and is avoided if possible. Only when the desirability of things which can be secured by labor is considered to outweigh the unpleasantness of mental or physical exertion will the effort be undertaken.

Mental effort is as truly labor as physical exertion. Probably the physician derives more pleasure from his work than a factory helper

does, but our only standard of whether a given effort is labor or not is whether it contributes to production. The efforts of the physician, the lawyer, the judge, the business executive, and the government official all add to the total amount of enjoyable goods and services which constitute the income of society. Some of it directly increases the flow of goods and services, some indirectly by making the work of others more productive.

Capital: Description. The third factor in production is *capital.* Ordinarily the term "capital" is synonymous with wealth; that is, useful material articles owned by man, but in discussing the factors of production it is convenient to limit the term "capital" to wealth which man has produced, thus excluding land. Land is distinguished from capital in this classification because of its peculiar relationship to production, but it should be remembered that it constitutes a particular and important class of capital in the ordinary meaning of that term. Capital, as we are using the term, falls into two major groupings. First there is an important group of capital articles which is desired primarily for its own sake rather than as instruments of production, including the mass of goods ready for consumption—food, clothing, furniture, etc. To this group we give the name *consumers' capital.* The second group consists of tools, implements, machines, and materials which are to be used to produce consumable goods and is called *intermediate goods* or *producers' capital.*

It is, however, not always possible to draw a sharp line between producers' and consumers' capital, and the terms are not always mutually exclusive. Some goods are capable either of being used for further production or of yielding their satisfactions at once. Coal may be mined for use in factories, or it may be used in the kitchen stove with the double purpose of warming the room and cooking food, or it may be burned in the open grate. A ship may be capable of use as a yacht or as a carrier of freight. Moreover virtually every sort of consumers' capital may, when used to satisfy wants, be at the same time serving the ends of production. This is obviously true of the form of consumers' capital *par excellence,* food, and it will be recognized as true of many of the less essential forms and even of many mere luxuries, such as athletic goods. But in spite of these qualifications we shall find the concepts useful in analysis.

Capital in production. We are at the present time so accustomed

to tools and machinery of every description that it is difficult to imagine conditions in which they were entirely lacking; yet at some stage, long before the beginnings of history, man must have been dependent on the unaided efforts of his hands. Such a situation, where no capital instruments are used, and brute strength or cunning is relied upon to secure the means of subsistence, may be called *direct production.*

What the first implement was is a matter of conjecture; it may have been nothing more highly developed than a club. Then possibly someone conceived the idea of attaching a sharp stone to the club, and an axe or hatchet was the result. Certain it is that tools came into use very slowly, and for centuries only the rudest and most primitive were known. Even during most of the centuries of which we have historical records, the machines and tools were of the simplest sort. But with the first tool came a definite advance, for man had started on his yet unfinished process of changing the conditions of his environment to meet his rapidly growing needs.

Production with tools and implements is called *indirect* or *roundabout* production. One may, for example, acquire the art of catching fish with his hands—this is direct production; but if he take time to make a fishing net he can greatly increase the effectiveness of his labor, including, as of course we must, the labor spent in making the net. Instead of proceeding directly to secure the desired products, time and energy are first spent in making tools, with the result that the production is made vastly more effective, and, passing beyond the small range of activities open to one without tools, production brings forth things which could not conceivably have been made with the unaided hands.

So apparent are the economies of indirect production that we need not be particularly concerned with the sharp contrast between direct and indirect methods of production. The contrast of significance is between a less and a more intensive resort to the roundabout methods of production. The labor necessary for the erection of a modern flour mill may be spread over many years, the materials for its construction may be gathered from the four corners of the earth, and hundreds of men may be engaged on one or another of the preliminary processes, whereas the old-fashioned gristmill could be made in a few weeks' time by a relatively small number of laborers; but the output of the modern mill per

unit of labor necessary to erect and operate it is immeasurably greater than came from its predecessor.

Wherever we find an example of intense recourse to the roundabout or capitalistic method of production, we may safely conclude that it is because the effectiveness of labor, as measured by the flow of goods, is thereby enhanced. Capital instruments are labor-saving devices, not only from an individual's point of view, but from the point of view of society as well.

The accumulation of capital. Saving. The degree to which indirect methods of production may be carried depends, in part at least, on the quantity of capital which is available for productive purposes. Other elements enter, such as the progress of invention, the size of the plants in the industry, the demand for handmade articles, and so on, but the quantity of capital is so fundamental a consideration that a study of the way in which it is accumulated will repay our consideration.

The formation of capital depends upon the readiness of men to postpone the gratification of wants in the present in the hope of increased satisfactions in the future; in other words, it rests on *saving*. In relatively simple cases this is quite clear. The savage who has gathered a supply of corn may be tempted to consume it before the spring, but if he can save enough for seed the following year he can hope for a more abundant supply in the autumn. A farmer today may use the proceeds from the sale of his crops to buy either a new automobile or an agricultural machine. The former will yield him pleasure at once, but the latter will enable him to expand his farming operations and to have a larger surplus of grain in the future. Nearly every manufacturer is compelled at some time or other in his career to choose whether to use the profits of the business to expand his scale of living or to "put them back in the business" in the form of enlarged facilities.

In these cases we can easily see that self-denial—saving—has added to the stock of capital. For the majority of people, however, the connection between saving and the productive processes is not so clear or so direct. Today the great mass of people obtain their living by exchanging services for money and with this money purchasing the various articles which are required or desired. To the individual the saving of a part of his income does not seem directly to affect production; it appeals to him solely because of the necessity of providing against illness or old age, or because

money can be left in a savings bank or loaned or used to purchase stocks and bonds and can thereby earn interest. Yet it has precisely the same influence on production as saving by the manufacturer.

If we consider saving from the standpoint of society as a whole, it will be apparent that it entails the devotion to the production of intermediate goods of effort which might have been spent in producing goods capable of yielding their satisfactions at once. It further implies that the consumption needs of those working on the intermediate goods are met from the flow of consumable goods produced by others. Were it not so they could not live.

We can imagine a simple kind of social organization where the tasks are arbitrarily divided among the different men and women, where all share in the common product, and where the farmers set aside a portion of their produce to support the miners and others working on intermediate goods, receiving their recompense eventually in the form of agricultural implements which will greatly increase the effectiveness of their efforts.

But although we may be able to find isolated situations where the intermediate worker is supported by direct payment in commodities, such as "grubstaking" a prospector, it is far from the normal case. The ordinary individual saves money, not commodities, and the laborer receives his remuneration in the form of money. This feature serves to obscure the nature of the problem and make it seem more difficult than it really is.

A clerk who saves $500 in the course of a year does so by denying himself the satisfactions which might have been obtained from the consumption of goods. When he places the sum in the savings bank, it is not kept in the bank's vaults until such time as he may require it but is loaned to some borrower. Now it is possible that the borrower may want the money to purchase goods for his own immediate satisfaction, but most loans are made to men actively engaged in business for business purposes. Let us suppose that the savings bank uses the money to buy a bond, a part of an issue of bonds sold by a railroad in order to be able to finance the construction of a branch line. The railroad uses the funds thus secured to engage men to work on the construction of the line and to purchase the results of past labor—steel rails, ties, etc.

The clerk has actually passed on his control over consumable goods to the laborers engaged in the intermediate processes; by reducing his own consumption he has made it possible for them to

work on nonconsumable goods and yet to live. During the years which may be required to construct the railroad line the men engaged in all the preparatory stages might have been producing food, clothing, etc., and for the time being have greatly increased the flow of consumable goods, but the eventual result of abstinence on the part of some in society is the creation of a great machine of production, the railroad, which makes the efforts of workers in all lines many times more fruitful. The saver gets his reward in the form of interest, but he also shares in the social gain, which consists of a more abundant flow of consumable goods.

It is difficult to overestimate the importance of the capitalistic or indirect methods of production and the necessity of saving, and it is easy to overestimate the adequacy of the volume of savings at present available in the world. Were the people of the nations of the earth suddenly to stop saving, spending all earnings on consumable goods, it would be but a short time before the entire structure of our civilization would be crumbling in the dust.

On a smaller scale this happened in Russia. In the days before the First World War Russia's industrial fabric depended in large measure on loans from foreign nations. Under the revolutionary and early soviet régimes Russia to a large extent was cut off from foreign sources of supply and in the disorder which prevailed was living in part on her capital. For a time there was practically no replacement of capital goods which had deteriorated or worn out. The result was an enormous decrease in the efficiency of Russian labor and the threatened extinction of many lines of industry.

The form of capital instruments. Invention. The effectiveness of labor in production depends not only upon the amount of capital available for use in production, but also upon the form which this capital takes. At the present stage in industrial development we could doubtless use to advantage a greater number of the tools and machines with which we are already familiar, but there must be some limit to the gain which comes from duplicating our present machinery. A more profitable line of development lies in improving the type of machines. Not two printing presses of the same capacity, but one improved press with twice the ordinary capacity.

The inventive genius of the world is constantly studying the possibilities of improving old machines and devising new machines to take care of processes now performed by hand. Society profits by the inventions which make it possible to secure a greater flow

of commodities with a smaller expenditure of effort. The saving which we have expressed in general terms appears to individual members of society in the guise of a lower price and therefore means that a given income represents command over a larger number of commodities than formerly. Measures which stimulate inventive genius are generally in the public interest. The inventor or the one to whom he transfers his invention may sometimes make a fortune in a few years through the monopoly which a patent right confers on him, but the people as a whole benefit still more. In the long run the hope of making a fortune is the spur which leads the inventor to persevere through the period of fruitless experimentation. The wealth which Edison accumulated appears a small price to pay for the services he performed for society in placing so many comforts within the reach of people of modest means.

The entrepreneur. Land, labor, and capital cannot work together without organization, direction, and control. Moreover in every productive undertaking there is *risk; i.e.,* the chance that the enterprise may prove unprofitable and that the capital devoted to it may be lost and the human effort wasted. The owners of the several factors of production are, as such, not prepared to undertake this planning function or to assume the risks of business. The landowner will provide land for a business if somebody will guarantee him a fixed rent. The laborer will work for an employer if the latter will by contract make himself legally liable to pay a definite wage. Finally the capitalist, in so far as he is performing the capitalistic function exclusively, lends money at interest to someone who can persuade him that the amount of the loan will be paid back in full.

Someone must borrow the capital, decide what to produce, and employ labor and land to produce it, in the hope that the venture will be successful enough to pay him profits over and above the fixed interest which he must pay to the owner of the capital. If the business is a failure, he must use his own resources to pay back the capitalist and to discharge his legal obligations to the landowner and the laborers. *Anyone who performs the function of organization, planning, and risk-bearing is termed an entrepreneur.*

The entrepreneur is thus the man who makes the fundamental decisions as to what commodities to produce and how to combine labor, land, and capital in their production. He determines the degree to which it is profitable to supplant manual labor by labor-

saving machinery, the amount of a given good which can profitably be produced. Whether a new invention is to be given a trial depends on his judgment. He studies the nature of the industrial operations, tries to simplify them as much as possible, and assigns laborers to one task or another according to his opinion of their abilities. The responsibility for the entire internal and external organization of a business falls on his shoulders—the coördination of production, finance, buying, and selling. The larger the business, the more incentive there is to delegate to others all tasks of a routine nature, and the more closely does the work of the chief executive partake of the nature of decision-making.

So important is this function in the modern economic organization that the entrepreneur is by some writers elevated to the rank of a fourth factor of production. As we have chosen to define the three factors of production, it is obvious that the entrepreneur is a human being, and whatever he does falls within our definition of labor. However, it would be equally correct to think in terms of four factors, thus setting off the services peculiar to the function of the entrepreneur from the other services of labor.

No corresponding class division. Whether we think of three factors or four, we must remember that this analysis of the factors of production does not imply a corresponding exclusive classification of human beings. There are landowners, and owners of capital, and laborers, and entrepreneurs; but these groups are not mutually exclusive. Landowners are frequently owners of capital, and capital owners—capitalists—very commonly own some land.

Many a laboring man is the owner of land, if no more than that on which his home is built. He is often also the possessor of at least some small amount of capital, in the form of stocks and bonds or a savings bank deposit. And the entrepreneur frequently owns some land and is generally a considerable owner of capital.

Division of labor: Coöperation. The key to an understanding of the complexities of present-day production is to be found in the idea of *coöperation.* By this is meant that the diverse elements in society are consciously or unconsciously working together toward a common end.

There are historical examples of communities working together and consciously regulating their various relationships to each other. Such societies as the Oneida Community, the Amana Society, Brook Farm, and others were organized for this purpose. These examples

of small isolated socialistic communities are, however, of only incidential interest to our present inquiry. Our task is to obtain a bird's-eye view of our present-day economic organization in order that we may see in what sense it is a coöperative system and may discover the motive forces which impel the various members of society to work together and the regulating forces which direct their efforts.

Probably the earliest and certainly the simplest type of coöperation is found where men get together to perform a task which is too great for the strength of a single man. In former years the farmers of New England worked together at a barn-raising or in lifting a millstone into place. There are and always will be in everyday life countless examples of this form of coöperation, which we may call *simple coöperation,* and which is distinguished by the fact that the several coöperating workers are all performing the same kind of work.

Division of labor. Coöperation becomes of vastly greater importance when it passes beyond this simple stage and leads to specialization, in which each worker confines himself to one kind of work, made possible for him because other kinds of work are cared for by other workers. Such coöperation is called *division of labor.* It is of varying complexity according to the degree of specialization.

The first steps occur when industry is separated into distinct trades, when one man devotes his entire time to making shoes, another to raising wheat, another to making cloth, and so on through the various trades with which we are familiar. But the mere separation of trades, in which each worker performs all the several operations of his own trade, taking the raw materials and finally turning out the finished product, has only begun to explore the possibilities of division of labor. Every trade is itself a complex of many distinct operations, and division of labor has not reached its final goal until each of these operations is made a separate task and assigned to a special group of workers. Although it did not originate there, it is in the modern factory that the more complex division of labor has been developed to its highest point. Here we see the extent to which it has been found possible to split a trade up into distinct minor operations, each performed by a different group of workers. In the making of a pair of shoes the soles are cut by one man, the various parts of the uppers by others, one part is stitched by still another, and so on.

Complex division of labor may be carried on without power machinery, but its possibilities are usually greatly enhanced by the employment of machinery for many of the operations. In the making of the Goodyear welt shoes it is said that there are 185 distinct operations, of which 157 are performed by machines.

Economies of division of labor. The predominant fact about division of labor is that it increases man's efficiency in production. In the first place, division of labor, by permitting each worker to specialize upon one trade or even upon a single operation in a trade, permits him to acquire a degree of skill and dexterity which would otherwise be out of the question. Constant repetition gives a manual dexterity which soon becomes automatic and almost unconscious. Just so far as it is necessary to think out the next move, speed is hampered. Furthermore the simplification of the individual operations in a factory makes it possible for a worker to become amazingly skillful in his given operation.

Division of labor also makes possible the saving of time which the unspecialized worker wastes in putting away the tools required for one job and getting the tools for another. A certain amount of time is used in really getting started, in deciding where and how to begin on the new job. When the day is divided among several different tasks, much time can be wasted in the motions of work without accomplishing a great deal.

When division of labor does not prevail, there is also the social waste which comes of leaving tools idle during a good part of the day. If analyzed, it can be seen that this is also a waste of time for the group as a whole. Assume that we have five individuals and five possible trades. If each individual follows all five trades, twenty-five sets of tools are necessary; if there is specialization, only five sets will be required, and they can certainly be made in a fraction of the time required to make the twenty-five, thus releasing labor for other employment.

Division of labor not only economizes the use of tools and machines, but it leads to the invention and utilization of complicated labor-saving machinery which would otherwise not be possible or economical. As the processes of any trade are subdivided each operation becomes more and more simple, and sooner or later it becomes possible to turn over the simpler ones to machines, thus setting free labor for other tasks and increasing the product of industry. The reduction in the physical labor of tending machines

has made it possible to use women for a variety of tasks which men alone used to perform, and the decreased difficulty of the task itself has considerably shortened the time spent in acquiring the necessary skill to perform it.

Finally with division of labor it becomes possible for a man to find the work for which he is best fitted and in which he can show the greatest dexterity and speed. It by no means follows that in our society this adaptation will always take place, and in many cases it is quite apparent that it does not take place. There is a tendency to follow customary lines of work. The son of a banker may follow banking as the line of least resistance, even though his natural bent may be in the direction of medicine. There is many an automobile mechanic who would be a more useful member of society as a farm hand, and doubtless many a farmer who is wasting his talents on the farm. But there is at least the possibility of this adaptation today, whereas formerly nearly every man had to be a farm laborer whether he would or not.

As a special application of this principle, we have the inestimable advantage of giving full-time employment to capable business leaders, who fill such an important place in our economic organization. If one will think of some of the great industrial leaders—Carnegie, Harriman, and Ford, to mention only three names—and try to imagine them in the medieval period in any rank of life he chooses, he can better appreciate the extent of the opportunity which the present organization of society gives to native talent and the degree to which society profits from the exercise of such talent.

All of these results of division of labor promote production by economizing labor, increasing its efficiency, and making more effective use of capital. With a given amount of human labor there emerges a richer flow of commodities and services for the satisfaction of human wants. Or we can look at it from another standpoint and say that division of labor makes it possible to turn out the same quantity of goods with less expenditure of effort, thus leaving more leisure for the enjoyment of the products of industry.

Effects on labor. It is sometimes objected that modern specialized labor is monotonous and deadening in its effects on the workers, that modern machine-tending is so largely automatic that skill is at a discount and labor is kept at a dead level, that the highly specialized laborer is unfitted for work in another trade. There is doubtless some measure of truth in these charges, but they may

easily be exaggerated. Labor prior to the development of modern industry was hardly free from monotony, and the shorter working hours of modern industry undoubtedly turn the scale in favor of the present day as regards the deadening effect of labor. For machine labor it can be said that, automatic though it be, the demand for skilled labor is still active and that labor could scarcely be more nearly on a dead level than it was during the precapitalistic era. The charge that specialized training today makes change of occupation difficult is the reverse of the truth. An operative in a watch factory could fit into a bicycle factory more easily than a house painter could become a blacksmith. On the whole such drawbacks as are inherent in the division of labor are certainly not sufficient to prevent its net effect being overwhelmingly on the credit side of the account.

Division of labor and the extent of the market. Whether the farmer builds his own barn in addition to farming or hires a carpenter to do it for him depends in part on conditions over which he, as an individual, has little or no control. Assuming that his true bent is farming, he will prefer to hire the carpenter, but if he cannot sell the wheat that he does not need for his own family, he will have to build his barn. If the carpenter cannot sell his services he is forced to raise his own wheat, even though he be an indifferent farmer.

The possibility of exchanging the surplus which one individual produces is a necessary preliminary to division of labor, and as exchange becomes easier there is a tendency to subdivide the tasks and to proceed from simple to complex division of labor. When the market consists only of the town and the countryside, the village cobbler has little use for assistance and can himself fabricate the whole shoe. If he does require one or two helpers, there will be little division of labor among them. Each will be able and will be required to make a whole shoe, although it may develop that one will be particularly skillful in a special part of the work and devote a large part of his time to that operation. But if the shoemaker is called upon to manufacture several thousand pairs of shoes a year, it will be economical for him to hire men to perform minor operations, for then he can provide them with continuous work. It would further be profitable for him to use machinery to perform some of the operations.

Territorial division of labor. The extent to which division of

labor will be carried depends then on the extent of the market, on the volume of goods which consumers are willing to purchase. The demand for a particular product may increase with the gradual growth of the population, as the result of a change in popular taste, or with the rise of average individual incomes. But in the past the most striking cause of increased demand for the product of a particular manufacturer or locality has been a widening of the market through improvements in the means of transportation; in other words, through the addition of a demand which was previously satisfied by other producers or not satisfied at all.

Certain regions are peculiarly fitted by natural endowment for the pursuance of certain occupations. The soil and climate of the American southern states make them particularly suitable for the growing of cotton; the coal and iron of Pennsylvania make specialization in iron and steel products natural; the fertile plains of the midwestern states offer unusual advantages for growing grain. The specialization in production which is thus dictated is called *territorial* or *regional division of labor.*

Under almost any circumstances there will be some degree of specialization in the products of a region, but unless adequate transportation facilities exist between different regions each must be self-reliant to a great degree. It is a familiar fact that a century ago New England was in large measure dependent on her own food supply, in spite of the fact that her soil has never been so productive as that of the West. Yet New England could not specialize in manufacturing until she could market a volume of goods far in excess of the needs of her own people and could satisfy her food requirements by grain from the West. The great farming regions of the West, on the other hand, would never have been able to specialize in agriculture except for the possibility of selling their products in New England and other regions and purchasing from New England and other regions their stock of manufactured goods. It was the development of the transportation system that made possible this interchange of goods and this territorial division of labor, to the great advantage of both regions.

Division of labor a form of coöperation. Production with division of labor is clearly coöperative. Laborers, employers, capitalists, and landowners are working together in the sense that each is in part dependent on the efforts of others. In the long chain of productive operations each producer is dependent on

the one below him for his raw materials. The cotton manufacturer relies on the grower of cotton, the cotton middleman, the coal miner, the manufacturer of cotton machinery, and literally a host of others, to provide him with the wherewithal to conduct his operations. The agriculturist grows more wheat than he can use, confident of his ability to sell the surplus and with the proceeds secure the cotton cloth that he needs.

Coöperation in production consists in the creation of a surplus and the exchange of that surplus for the products of others. The laborer working for wages just as truly exchanges his surplus as does the manufacturer who hires him, for the manufacturer pays him the present worth of his product in money instead of giving him a portion of the product to sell, as might be done and as is sometimes done with farm labor.

This fact of coöperation is by no means of academic interest only; it is of great practical importance. In the past, material progress has followed in the steps of coöperation. As in the past so in the future, the path of material progress lies in the direction of increased coöperation, of greater specialization both within industry and between regions having natural aptitudes in the production of certain goods. Any artificial restraint therefore which hinders this specialization decreases by so much the effectiveness of labor.

EXERCISES

1. The creation of a physical object may be unproductive; and there may be production when no physical object is created. Explain what is meant, and give illustrations.
2. Which of the following are producers? Explain in each case.
 (*a*) A stenographer. (*b*) A locomotive engineer. (*c*) A life insurance salesman. (*d*) A racketeer. (*e*) A dentist. (*f*) A mechanic. (*g*) A beggar. (*h*) A burglar. (*i*) A professional baseball player. (*j*) A fortune teller.
3. A, B, C, D, E, F, G, and H each save $1,000. A buys a piece of land. B deposits his savings in a bank. C invests his in the stock of a newly formed paper-manufacturing enterprise. D spends his on a trip to Hawaii. E sets up an automobile repair shop. F places a $1,000 bill in his safety deposit box. G buys a $1,000 bond of the U. S. Government. H buys $1,000 worth of stock of the American Telephone and Telegraph Corporation. Explain how each of these activities affects the formation of capital.
4. A producer of shoes makes a profit of $10,000. He takes $5,000 for his current living expenses, leaving the rest "in the business." Explain how and to what extent this affects the formation of capital.

5. Show what factors of production are combined in each of the following enterprises:
 (*a*) A dairy farm. (*b*) A coal mine. (*c*) A retail store. (*d*) A doctor's office. (*e*) A fishing enterprise.
6. In each of the enterprises of exercise 5 above, identify the entrepreneur and describe his contribution to the process of production.

CHAPTER III

The Business Firm

Function of the business firm. In the course of the extraordinary era of economic progress ushered in by the Industrial Revolution, old ways of conducting business have been modified, and new forms of business organization have been developed, thus enabling the several branches of industry to adapt themselves to changing conditions and to function more easily and efficiently. Under the conditions of economic freedom that generally prevailed, the hope of increased profits offered the compelling motive back of this evolution.

In the modern free society the business firm is the device used by the entrepreneur to bring together the factors of production and employ them in producing. We may think of a business firm as a combination or collection of items of land, labor, and capital, assembled by an entrepreneur for the purpose of making the maximum profit possible through the process of producing and selling goods and services to satisfy human wants.

To obtain the use of the land, labor, and capital required for production, the entrepreneur has to offer inducements which will be attractive to those persons who own or control these factors. The bringing together of the resources and talents of a large number of persons into a coöperative effort in production thus gives rise to what may be called the legal problem of establishing and defining the rights and obligations of each of the participants in the coöperative endeavor.

Each of the several forms of the business firm involves, therefore, a particular set of property rights or legally recognized arrangements under which individuals may combine their resources and talents in seeking profits and by which is established the participation of each person in the amount of the income earned and in the control and risks of the business. From this point of view, the business firm is essentially the expression of a group of ideas or mutually acceptable arrangements under which individuals are

allowed to work together in producing and to act effectively and promptly as a group in the multitudinous transactions which business involves.

The individual proprietorship. The simplest form of business firm is the individual proprietorship. In this form of organization the proprietor is in sole charge of the business, responsible for its success or its failure. He may (in most states) start up in business when he chooses and cease operations when he so desires, without the bother and expense of legal formalities. Unless an activity is specifically prohibited by law, such as the coining of money or the carrying of mail, no line of business is closed to him. Responsible to himself alone, he is in position to make decisions at once without seeking either the advice or the assent of an associate, and he is thus able to take advantage of the opportunity of the moment. His business secrets cannot become common property through the indiscretion of associates.

While there are advantages for the small business in this form of organization, there are certain drawbacks which make it undesirable for a concern of any great size. In the first place the individual is seldom able to invest as much capital in his own business as can be secured by a partnership or a corporation. And if he is able to invest a large amount of capital, the great risk involved is a deterrent, for he is personally liable for all debts of his business.

Hence at the present time we find relatively few large enterprises organized and operated by a single individual. While the number of manufacturing plants owned and operated by individuals is over one-half of the total number, the output of these plants forms an insignificant part of the total output. In agriculture alone the individual is still the predominant figure.

The partnership. The partnership is a more complex type of organization than the individual proprietorship, for it involves relations between the individual partners as well as the relationship of the partnership to the public. Essentially it is a voluntary association of two or more individuals for the performance of a specific or a general object. This object may be accomplished in a day or a week, or in any longer period, or the association may be formed for more general purposes, such as the establishment and operation of a factory.

In any case the purposes for which the partnership is formed, its

duration, the ways in which it may be dissolved, and the mutual rights and obligations of each partner to the other will be the subject of an agreement called a *contract,* usually, but not necessarily, set forth in written form. As was true of the individual proprietorship, the partnership can be created with little legal formality and without delay.

While the partnership agreement determines the relations between the individual partners, it cannot control the relations of the individual partners, either singly or as a group, with the general public. Legally, each partner is an agent of the partnership, fully empowered to bind the group in all lawful undertakings reasonably within the scope of the firm's business. Each partner is thus said to be a *general agent* for the firm.

In the event of inability to meet all debts from the funds of the partnership, each partner is liable to the full extent of his personal resources for the debts of the business, even though the agreement may make one partner responsible for only a certain proportion. If Jones and Brown suffer large losses and the debts exceed the resources of the partnership, and Jones finds himself unable to meet his share of the losses, Brown must make up the difference if he is able to do so. This is known as *unlimited liability* and is one of the chief points of difference between the partnership and the corporation.[1]

The fact that there are several men united together makes it possible for the partnership to control a larger amount of capital than can the individual. Few concerns are able to finance themselves with their own capital exclusively; they must rely to some extent on bank credit, secured through loans from banks, and on commercial credit, obtained by purchasing goods to be paid for in the future. The ability of the partnership to secure capital in these ways is likely to be greater than that of the individual, in so far as the unlimited liability of several men offers a greater margin of safety to the creditor than the unlimited liability of one.

Another advantage of the partnership over the individual proprietorship may come from the association of abilities. An important question will be viewed from more than one point of view, and the decision which is reached may therefore be sounder. Offsetting this there is the possibility of serious disagreement and

[1] It should be noted that there are a few limited liability partnerships, which are in this respect similar to the corporation. These are exceptional.

consequent delay in making decisions, but at least snap judgments are likely to be avoided. A certain degree of specialization is possible, as in a wholesale house where one partner is entrusted with the buying, another with the selling, and the third with financial questions.

One of the chief weaknesses of the partnership is the prospect of its enforced dissolution on the death of a partner. Under the common law doctrine,[1] unless modified by statute law, the death of one of the partners automatically dissolves the partnership and makes necessary an accounting and distribution to his heirs of the deceased partner's share of the net resources of the partnership. This may be a hardship, both to the partnership and to the heirs. But the heirs cannot join the partnership without the formation of a new partnership.

The theory on which this doctrine is based is that the partnership is a peculiarly personal relationship and that anything which occurs to interfere with this must necessarily bring the original arrangement to an end. A logical corollary is that a partner may not transfer his interest in the partnership without causing a dissolution of the partnership, because no associate can be forced on the other partners against their will.

But the partnership has such obvious advantages over the individual proprietorship that, before the corporate form was devised and made common, it was the prevailing form of organization in enterprises of moderate and large size, and it is still not uncommon to find large concerns conducted for many years under a partnership agreement, although the corporation has steadily increased in favor and now overshadows the partnership in importance.

The corporation: Modern growth. Though the most significant type of business unit in the United States at the present time, the *corporation* is of comparatively recent development, the great increase in the number of corporations having come since the Civil War. In the fields of railroading, banking, and insurance the corporation is practically supreme today.

Even in manufacturing, where its growth has been slower, it has a position of dominance. In 1939, although only 52 per cent of all

[1] Common law is based on universal custom which has received recognition and sanction in decisions rendered by the courts. Statute law consists of laws enacted by a legislative body, such as the Congress of the United States or the legislature of one of the states, and enrolled on the statute books.

manufacturing establishments in the United States were owned by corporations, they turned out almost 93 per cent of the total value of products, and employed nearly 90 per cent of all wage earners in manufacturing. In some lines of manufacturing the importance of the corporation is even more striking. For example, in the manufacture of rubber boots and shoes, linoleum, flat glass, rayon, and steam engines, corporations have entire control. They turn out nearly 99 per cent of all the agricultural implements and of the clocks and watches made in the United States.

In occupations formerly thought the province of the small retailer, the corporation has made its appearance, as is evidenced by the rapid growth of chain stores in the grocery, drug, tobacco, restaurant, and other fields. Taking the nation's activity as a whole, 57 per cent of the total American national income in 1929 was produced under the corporate form of business enterprise and the remaining 43 per cent under other forms of business organization.[1]

Despite the continued growth in the number and importance of corporations, it must not be thought that they have supplanted other forms of business organization. In certain lines of industry such as agriculture, where the corporation is rarely found, and the building industry, the partnership or the individual proprietorship is still either the dominant or an important form of organization.

Nature of the corporation. A corporation has been defined in part as "an artificial being, invisible, intangible, and existing only in contemplation of law." This definition indicates one of the most distinctive characteristics of the corporation; in the eyes of the law it is a legal or artificial being, a fictitious person, conceived to have an existence apart from that of its owners. By contrast, the partnership exists in the persons of the partners. For example, suits by or against the partnership are brought in the names of the partners, but the corporation sues and is sued in the corporate name.

Furthermore the corporation depends for its existence upon the sanction of the law. The individual businessman and the partnership carry on business by virtue of rights long recognized by common law, but the corporation exists only through the official act of the state. If the organizers fulfill the requirements as laid down in the corporation laws and make application in due form

[1] Twentieth Century Fund, Inc., *Big Business: Its Growth and Its Place.* New York: The Fund, 1937, p. 17.

to the proper state official, a *charter,* or *certificate of incorporation,* will be issued to them defining the rights and duties of the corporation.

Called into being by the act of the state, the corporation exercises only such powers as have been conferred upon it. A partnership may engage in any activity which is not prohibited, but the corporation is restricted to the range of activities sanctioned by the terms of its charter or the general laws, either explicitly or implicitly, as essential to its existence. Acts beyond the scope of the corporation's powers are termed *ultra vires* and may result in the revocation of the charter and the dissolution of the corporation.

The doctrine of personal relationship, which is stressed in partnerships, is not so important in the case of corporations; the association of capital is of greater significance than the association of individuals, and this finds its exemplification in the right of each of the owners in the corporation—*i.e.,* each shareholder—to transfer his property rights in the corporation at will, without seeking the assent of the other owners, either individually or collectively. It follows that the death or disability of one of the owners (shareholders) in the corporation does not bring about its dissolution. This makes the corporation a more permanent form of organization.

Another distinctive feature of the corporation, as compared with the partnership, is the *limited liability* of the owners in the corporation for corporate debts. This is not a necessary mark of the corporation; there have been corporations the shareholders of which have had full personal liability for corporate debts, but it is usual at the present time to limit the liability of the owner to his investment in the business. A man who has purchased 500 shares of stock in a corporation is liable to lose what he paid for them, but the law does not lay hold of his personal resources to satisfy the creditors of the corporation, except in certain special cases where the law provides for *double liability,* as it formerly did in the case of stock issued by national banks. In this instance the owner of a $100 share of stock in a national bank was not only subject to the risk of losing the money he had invested but could also be required to contribute $100 more to meet the debts of the bank.

Finally we should note that the conduct of the affairs of the corporation is delegated by the owners to hired managers. The

shareholders elect representatives from among themselves, called *directors,* to care for their interests and to supervise in a general way the operation of the business, but the active management of the corporation falls upon the officers, who are elected by the directors. The board of directors is the policy-forming body. The stockholders do not participate directly in policy-forming or in the routine conduct of the business, though there are certain actions of the directors which law and custom require the stockholders to approve.

Capital stock. A corporation's capital stock represents ordinarily the original and permanent investment of the owners in the corporation. Suppose that a group of men incorporate the Steel Manufacturing Corporation and agree to contribute $1,000,000, divided into 10,000 shares of $100 each. A takes, let us say, 5,000 shares, and pays $500,000 into the treasury for them; B takes 1,000 shares, and so on until all the shares are sold. The corporation now owns $1,000,000 in cash, and the organizers of the corporation have certain property rights in the corporation.

The collective name for these property rights is *capital stock,* and if we say that the capital stock of the company is $1,000,000, we mean that $1,000,000 is the stated amount of the capital stock issued to the stockholders and that this is the accounting estimate of the amount which has been contributed by them to the firm.

Each owner of a share of stock, or *shareholder,* becomes a part owner in the corporation, and as such is possessed of certain rights in the corporation, such as the right to receive dividends when legally declared, the right to share in the resources of the company if it is liquidated, and usually the right to vote at the meetings of the shareholders.

A *certificate of stock,* signed by the proper officials of the corporation, serves both as evidence of the shareholder's rights of ownership and, when properly endorsed, as a convenient method of transferring his rights to another. It must be emphasized that a stock certificate is not synonymous with a share of stock; a share is one of the units into which the total property rights of the owners have been divided, and the certificate is merely presumptive evidence that a part of the total rights is vested in the holder. This is an example of the distinction between a property right and the document in evidence thereof, which was pointed out in Chapter I.

Many corporations find that their financial needs are best met by the issue of two types of capital stock, which differ from one another in some important respects. These are *common stock* and *preferred stock*. Common stock usually carries with it the exclusive right of voting at shareholders' meetings. Preferred stock also may confer the right of voting, but usually the preferred stockholder has no voice in the management so long as his dividends are regularly paid. The ownership therefore of more than one-half of the outstanding common stock will give control of the corporation, since the owner has one vote for each share of stock he holds. In fact in a large corporation a much smaller percentage will usually suffice because of the lack of interest on the part of a number of the small stockholders and of division among the larger ones.

Dividends are declared on common stock by action of the board of directors when profits have been earned and when it is judged wise to distribute them, and the amount or the rate of the dividend is entirely within the discretion of the board. Hence the stockholders in a new corporation are uncertain of a return on their investment for some time, conservative policy dictating that the earnings be reinvested in the business.

If the corporation succeeds in establishing a sound business, its dividend rate may be stabilized, and if it should prove especially successful it might have a dividend rate much higher than the rate of yield on any other type of security.

Preferred stock differs from common stock in that a fixed rate of return must be paid as a condition of any dividend payment on the common stock. Accordingly the rate of return is somewhat lower than that which can be expected eventually, and as a general rule, from common stock.

A preferred stock is said to be *noncumulative* when the preferred stockholder has no redress in case the company is unable to pay dividends in any given year; *i.e.*, even though during the succeeding year large profits are made, he does not then receive dividends for the previous year.

But preferred stock is more frequently *cumulative*. Then all current and back dividends on this stock must be paid before any dividends can be declared on the common stock. This protects the preferred stockholder against manipulation of the accounts to conceal profits and avoid thereby the paying of dividends.

Preferred stock may also be *participating*. This means that when

a dividend has been paid to the common stockholders at the same rate as to the preferred stockholders, both the common and the preferred stockholders share, in proportions already determined, in any further dividend payments made for that dividend period.

The owner of this type of stock has preference over the common stockholder in another direction also; *i.e.,* in the distribution of assets in case of liquidation. The common stockholder receives his share of the assets only after the preferred stockholders have been paid in full. If the remaining assets are insufficient to redeem the common stock in full, the amount is divided pro rata among the common stockholders; if there is more than enough to redeem the common stock, then the remainder is usually divided pro rata among the common stockholders.

Bonds and notes. Another part of the funds for the operation of the corporation is often secured through the sale of its bonds and notes. A bond is a certificate of indebtedness issued by a corporation, bearing interest at a stipulated rate and payable at stated times, and having a certain number of years to run before the corporation is obliged to redeem it. Essentially it differs from a promissory note only in the length of time before maturity, the type of security on which it rests, and the greater formality with which it is issued.

The notes of a corporation, apart from the promissory notes of a few months' duration, partake of most of the characteristics of the bond. The majority of them however run for a few years only, whereas bonds may mature many years after their issue. The notes frequently pay a higher rate of interest than do the long-time bonds. This arises from the fact that note issues are often resorted to at a time when the ruling rate of interest is too high to make it desirable to issue bonds; by the time the notes mature it is expected that the interest rate will have fallen so that, if necessary, long-time bonds may be sold to replace them.

The bondholder has no voice in the management of the business so long as his interest payments are met with regularity, except in so far as his contract with the corporation may limit its freedom of action in respect of matters which have a bearing on its ability to pay the principal or interest of the bond. If the security behind the bonds is a mortgage on the physical equipment of the concern, as is frequently true, default in the payment of either interest or principal may force the bondholders to foreclose the mortgage and

take over the operation of the business to protect their own interests. In the liquidation of a business the bonds are redeemed before either preferred or common stockholders receive anything.

Stocks and bonds compared. The essential distinction between stocks and bonds is that the latter represent a definite contract on the part of the corporation to pay money to the bondholders, whereas stocks only entitle their holders to receive their shares of such business profits as may be earned and their shares of the corporation's net assets upon liquidation.

As regards the risk of failure to receive payment of income and return of the amount invested, the bonds of any given corporation are obviously safer than its stocks. On the other hand, they carry no chance of returns beyond the fixed interest rate, and they give no control over the conduct of the corporation's business. Preferred stocks carry more risk than bonds but give the chance of a somewhat larger return; they generally give no control. The common stockholders take the greatest risk but enjoy the chance of the highest return; and they have usually complete control of the corporation's management.

Of course this does not mean that bonds are always safer than stocks. The stocks of a sound, well-managed corporation would carry less risk of loss than the bonds of a corporation whose assets were of questionable value or whose management was inefficient or dishonest.

There is however another risk to which all investors are subject; *i.e.*, the risk of a decline in the purchasing power of the monetary unit. Here the bondholder bears the greater risk. Both his income and his principal are defined in the contract with the corporation in terms of dollars. If the general level of prices should rise, the bondholder, with the fixed number of dollars received as interest from each of his bonds, could buy a smaller quantity of commodities. The shareholder, either by an increase in the dividend rate or through an appreciation in the market value of his securities, may be left in a financial position substantially unaltered or even improved. By the same token, the bondholder stands to gain from a fall in the general level of prices which is not violent enough to threaten the soundness of the business. The stockholder gambles more than the bondholder on the success of the enterprise. The bondholder gambles more than the stockholder on the price level.

The property account of a corporation: Assets and liabilities. Many of the features of fundamental importance in connection with the corporation can be illustrated to best advantage by the use of a simple *property account,* or *balance sheet.*

A balance sheet is a record, as of a particular time, of a real or fictitious person's *assets* and *liabilities* and of the relationship between them. Assets consist of wealth and property rights against other free persons and their wealth; a person's liabilities are property rights of others against him or his wealth. Thus if A owns a house, it appears on the asset side of his account; his promissory note to B for $500 appears on the liability side of his account and on the asset side of B's account.

Referring to our former example we will assume that the Steel Manufacturing Corporation has been incorporated with a capital stock of $1,000,000, all of which has been paid for in cash. The balance sheet would then appear as follows:

BALANCE SHEET OF STEEL MANUFACTURING CORPORATION

July 1, 1937

Assets		*Liabilities*	
Cash	$1,000,000	Capital Stock	$1,000,000

Before there can be a balance sheet, a fund of assets must have been segregated under a common ownership and control. These assets are owned by the corporation. Thus cash appears on the assets side of the balance sheet above. The total of claims against the corporation and through the corporation against its assets must be equal to the total fund of assets. These claims are called *equities* or *liabilities.* The corporation's capital stock is such an equity, since it represents the stockholders' claims against the corporation; it therefore appears on the liability or equity side of the balance sheet.

Now let us assume that the corporation markets $500,000 worth of its bonds, and spends $1,200,000 for plant, including machinery, etc. Its balance sheet would then read as shown in the balance sheet at the top of the following page.

We have now two types of liabilities—the capital stock, which is a property right of the owners, and the bonds, which are property rights of outsiders or creditors.

BALANCE SHEET OF STEEL MANUFACTURING CORPORATION

July 31, 1937

Assets		*Liabilities*	
Cash	$ 300,000	Bonds	$ 500,000
Plant	1,200,000	Capital Stock	1,000,000
	$1,500,000		$1,500,000

During the next ten years the plant is in operation making and selling the goods for which it was built. It has sold some of its goods and has purchased some of its raw materials on credit; it has on hand a supply of materials of various kinds, raw and finished, and it has increased its investment in the plant. Furthermore the years have been profitable, and we find all this reflected in the balance sheet of July 1, 1947, below.

BALANCE SHEET OF STEEL MANUFACTURING CORPORATION

July 1, 1947

Assets		*Liabilities*	
Cash	$ 102,560	Accounts Payable	$ 60,000
Accounts Receivable	75,000	Notes Payable	50,000
Notes Receivable	25,000	Interest due Bondholders	15,000
Raw Materials	100,000	Bonds	500,000
Finished Goods	150,000	Capital Stock	1,000,000
Plant	1,800,000	Surplus	627,560
	$2,252,560		$2,252,560

Some of the items on this balance sheet need explanation. The items "accounts receivable" and "notes receivable" among the assets are property rights which the corporation has against other persons and which are evidenced respectively by entries on the books of the corporation and by promissory notes. Presumably they arose from the sale of goods. Corresponding items among the liabilities refer to similar obligations to others on the part of this corporation.

The other item which needs explanation is the surplus. It will be noted that since the date of the last statement our assets have grown from $1,500,000 to $2,252,560, an increase of $752,560. At the same time claims of outsiders against the corporation have grown from $500,000 to a total of $625,000 (bonds, interest due, notes payable, and accounts payable), an increase of $125,000. The

increase in total assets exceeds the increase in liabilities to outsiders by $627,560. This represents a gain from all the corporation's business, and this gain belongs to the stockholders.

The corporation can never have assets greater than its liabilities; they must always balance. To indicate the growth in the value of the stockowners' holdings we might possibly add the $627,560 to the capital stock, but it is desirable to leave this item intact, in part to show the amount of the original investment, and partly for technical reasons. This new item is therefore introduced to represent the growth in the owners' share in the assets of the corporation.

It should be emphasized that surplus is not cash or money or any form of wealth. It is a property right. The sum of capital stock and surplus gives the accountants' estimate of the total worth of the property rights of the owners against the corporation and is known as the *owners' equity* in the corporation, or the *net worth* of the corporation to the owners.[1] The net worth may always be found by subtracting the liabilities to all others than the owners from the total of the assets. In any balance sheet, the asset side is a list of property rights against wealth and persons, whereas the liability or equity side shows to whom these property rights as a whole belong. The liability side is thus a record of ownership.

Losses and insolvency. The relationship between these items representing ownership may perhaps be made clearer by taking a case of declining assets. Referring again to the balance sheet of July 1, 1947 (on page 46), let us suppose that a note for $10,000 proves to be uncollectible, making it necessary to reduce the notes receivable and the total assets by $10,000. Losses fall first upon the owners of the business, and we will therefore reduce the surplus by $10,000 to make the two sides balance. A fire occurs and wipes out the plant entirely, causing a loss of $500,000 in excess of the insurance. The total assets will be decreased by $500,000, and an-

[1] The word "capital" is a much abused term, used in various senses which have become so entrenched in popular and even scientific language that we, unfortunately cannot restrict it to one meaning without too violent a break with common usage. We have already defined capital as one of the factors of production (see Chapter II). We must now note that the term "capital" may be used also to designate what is in the text called the "net worth" of a corporation or of a partnership or individual. This should always be distinguished from the "capital stock" when it is not the same as the net worth. And finally we may use the same overworked term to designate all the resources at the disposal of a business enterprise, incorporated or not; that is, the total assets.

other $500,000 will be subtracted from the surplus and from the total liabilities or equities.

Losses can continue to occur without affecting the outside creditors' position until the point is reached where capital stock and surplus have been wiped out entirely. When that point is passed, the corporation is said to be insolvent. In precise terms a corporation is insolvent when its assets are less than its liabilities to others than the owners. When the assets are equal to or greater than the liabilities to outsiders, the corporation is solvent.

The value of stock. It has been customary to assign a definite value to each share of stock and to engrave this value on the stock certificate. This value is known as the *par value,* and the par value, multiplied by the number of shares outstanding, gives the figure at which the capital stock is carried on the books of the corporation.

While there are certain conveniences which attach to this custom, it has tended to confuse certain investors and others as to the true worth of a share of stock as reflected in either its actual or potential earning capacity. To engrave on a certificate of stock the words "Shares, $100 each" may be but the expression of a pious hope; the shares may be worth no more than the paper on which they are printed, or they may be worth much more than $100 per share. The increasing use of stock with no par value assigned is evidence of the recognition of this fact.

A somewhat more reliable index of real worth than par value is found in the *book value* of each share of stock. This is easily obtained by taking the amount of the owners' equity as shown by the balance sheet and dividing it by the number of shares. According to the last balance sheet above, the owners' equity was $1,627,560, and there were 10,000 shares outstanding; each share must then have a book value of $162.756. The book value fluctuates with the increase or decrease in the owners' equity; large earnings swell the surplus and send the book value up, while losses, which of course must be borne by the owners, decrease the surplus and cause book value to decrease. Since the asset values on the balance sheet consist predominantly of historical costs, rather than the present values, the book value per share of stock does not purport to represent its real, or market, value.

A third concept is the *market value* of a share of stock, or its selling price in the open market. This is based on the estimate

of the dealers in the market of the earning power of the corporation. If there is an active market for the stock of a particular corporation, enough interest will be manifested by investors and professional traders in the affairs of the corporation so that the market value, aside from temporary speculative fluctuation in either direction, may be taken as a fairly trustworthy indication of its true value. It represents the appraisal of those who may be presumed most competent to judge.

Interest payments and dividends. We have already stated that the payment of interest on bonds is a part of the contractual relation between the corporation and the bondholder. If we assume the interest rate in our example above to be 6 per cent per annum, payable semiannually, the corporation evidently has a charge of $15,000 to meet every six months. The payment of $15,000 for this purpose would change our balance sheet as follows: (1) cash would be decreased by $15,000, and (2) the item "interest due bondholders" would be wiped out entirely.[1] A simultaneous and equal change in both assets and equities leaves the totals equal, and no further change in the statement is required.

Let us assume next that the directors decide to pay a 5 per cent dividend in cash to the stockholders. This means 5 per cent on the capital stock, which is $50,000 or $5.00 per share of stock. Cash would be decreased by $50,000, and surplus would be similarly decreased. The owners' equity is reduced to $1,577,560, and if we calculate the book value of each share of stock we will see that it is now $157.756 (*i.e.*, $1,577,560 divided by 10,000); in other words it has decreased by precisely the amount paid in dividends on each share of stock.

Stock dividends. Suppose the Steel Manufacturing Corporation decides to declare a *stock dividend* of 50 per cent. Each shareholder will receive half as many shares as he now possesses. A, who owns 5,000 shares, will receive 2,500, and the other shareholders in proportion. Clearly this does not disturb the assets in the slightest; so far as our statement is concerned all the changes will be among the

[1] From the fact that the item "interest due bondholders" in the statement of July 1, 1947, is $15,000, the exact amount of the semiannual interest charge, we infer that this is the date on which the semiannual payment of interest is due. On any other date this item would be something less. Immediately after the payment it is zero, as stated in the text; thereafter it increases gradually, as the liability "accrues," until it again reaches $15,000 on the date when payment is due six months later.

equities. Capital stock will be increased to $1,500,000 and the surplus of $627,560 will be reduced to $127,560. The net worth of the business is not altered, but the book value of each share of stock is decreased, for now we divide the net worth by 15,000 instead of by 10,000.

It is quite evident that the position of the stockholders has not changed at all so far as their claims upon the assets of the corporation are concerned. The 5,000 shares which A owned had a book value before the stock dividend of $162.756 per share, or a total value of $813,780; now he holds 7,500 shares, each having a book value of $108.504, or a total value of $813,780. The book value per share of stock has declined, but the total book value is not affected. In fact the stock dividend, except in a figurative sense, is not a dividend at all, but merely a transfer from surplus to capital stock.

Since the stock dividend makes no change in the position of the stockholder with respect to the corporation, it may be asked what purpose it can serve. Without attempting a complete answer, the two most common purposes may be cited.

In the first place, it may happen that the market value of the stock is so far in advance of the par value that a share of stock is less easily traded in. Five shares which sell for $100 each are much more convenient than one share which sells for $500. The stock dividend may then be resorted to in order to increase the number of the shares and lower the value of each. But the chief motive is probably to avoid the appearance of unusually high profits and high dividend rates. A dividend of 20 per cent on a capital stock of $1,000,000 is no greater than a dividend of 10 per cent on a capital stock of $2,000,000, but it seems greater; the former may attract considerable notoriety, lead to the suspicion of monopoly, and invite competition, while the other passes by without exciting comment.

Watered stock. Stock is said to be "watered" when the corporation issuing it has not received full value for it. There are several different ways of watering the stock. One of the most common is to issue stock in payment for property or services which are overvalued. This may be done innocently and in good faith, or it may be done with intent to defraud. Suppose that $100,000 worth of stock, par value $100 per share, is sold to investors for cash, and that another block of $10[redacted] is given in payment for a tract of

land which, at the most liberal estimate, is worth not more than $50,000. The statement would show assets of $200,000 and liabilities of $200,000, and the book value of each share would be $100. A true record on the books of the corporation of the value of the assets would, however, record them as worth $150,000 and would thus show a book value of only $75 per share.

A stock dividend without a sufficient surplus to justify it is a form of stock-watering. In our example on page 49 a stock dividend of $1,000,000 could have been declared only by raising the recorded value of the assets by $372,440.

The essence of stock-watering is overvaluation of the assets of the corporation. Like the legitimate stock dividend, its purpose may be to disguise large profits, though in this case it really conceals them instead of merely avoiding an appearance of high profits. Whatever its purpose, stock-watering involves falsification of accounts and carries at least the presumption of dishonest intent.

The corporation's income account. The balance sheet or property account of a corporation shows the condition of business at a given instant of time, but it does not give all the information that may be required regarding the condition of the business. Even a comparison of two such accounts for successive years, while more informing than a single balance sheet, does not tell the whole story. It does indeed show the changes in the various items in the balance sheet, and it may shed some light on the profitableness of the business, but it does not tell us how the changes have come about, nor the amount of the profits or the losses, nor what has occasioned them. To complete the picture we must have another account, called the *income account* or the *profit and loss statement,* which gives in condensed form the history of the operations of the business over the accounting period in question.

The first goal of the profit and loss statement is to record the net profit from the operation of the business. For purposes of illustration, let us assume that the City Grocery Company is dealing in eggs solely. Its sales of eggs in the year 1946 amounted to $215,000. On January 1, it had on hand eggs which cost it $40,000. During the year it purchased eggs costing $171,200, and it paid out in freight and cartage $3,100 to bring these eggs to its warehouses. At the close of the year it had on hand eggs which had cost $42,300. From these figures we can compute the cost of goods sold, as follows:

Inventory Jan. 1, 1946	$ 40,000
Purchases	171,200
Freight and cartage inward	3,100
	$214,300
Less: Inventory Dec. 31, 1946	42,300
Cost of goods sold	$172,000

In addition to the cost of goods sold, there are other expenses which have been made and must be deducted. Expenditures for advertising, salesmen's salaries, commissions, and the maintenance of a delivery service are necessary to keep the merchandise moving from the warehouse to the consumer. Expenditures of this type are classified as *selling expenses* and must be deducted from gross profit. They cannot be included in the cost of goods sold, because, for the most part, they do not relate to any particular sale. Commissions could conceivably be deducted from selling prices, but convenience and uniformity demand that they be grouped with the other selling expenses.

In seeking the final net profit from operation it is also necessary to take into account another major group of expenses known as *general and administrative expense.* Such are salaries to executives and to the office employees who keep the books and accounts, send out monthly statements, make collections, adjust complaints, supervise the purchasing of merchandise, etc. Office supplies of all kinds are consumed. Moreover, provision must be made for warehouses, offices, showrooms, etc., with all the necessary equipment, and depreciation on all these assets must be treated as an expense. Characteristic of these expenses is the fact that they relate almost entirely to periods of time rather than to the volume of sales. Depreciation occurs, and rent and taxes accrue at a uniform rate whether sales be large or small.

We are now ready to set up the complete statement showing the *net profit from operation* or *net profit on sales.*

Logically the profit and loss statement should stop at the net operating profit, but in every business there are incidental receipts and payments, usually financial in character and not a part of the principal business of the corporation. Space may be leased to an outside tenant; interest may be received on investments; rent or interest may be paid; discounts may be earned or given, and the like. For want of a better title such income is called *other income* and

such expenses *other expense*. Adjusting the net operating profit for these additional items gives the *net profit* for the year, as below.

PROFIT AND LOSS STATEMENT OF THE CITY GROCERY COMPANY
(For the year ending December 31, 1946)

Net Sales			$215,000
Operating Expenses:			
Cost of Goods Sold		$172,000	
Selling Expense:			
Advertising	$ 5,000		
Salesmen's salaries	10,000		
Delivery expense	3,000	$ 18,000	
General and Administrative Expense:			
Salaries	$11,200		
Office expense	3,240		
General expense	2,000		
Depreciation	1,500		
Taxes	800	$18,740	$208,740
Net Operating Profit			$ 6,260
Add: Other Income:			
Interest earned		$ 300	
Rental income		240	540
			$ 6,800
Deduct: Other Expense:			
Interest paid		$ 1,100	
Loss on sale of real estate		700	1,800
Net Profit			$ 5,000
Dividend declared			$ 3,000
Net addition to surplus			2,000
Surplus, December 31, 1945			13,500
December 31, 1946			$ 15,500

A comparative balance sheet for the beginning and ending of the accounting period is given below.

COMPARATIVE BALANCE SHEET OF THE CITY GROCERY COMPANY
(As of the close of business December 31, 1945, and December 31, 1946)

Assets	Dec. 31, 1945	Dec. 31, 1946
Cash	$ 5,000	$ 4,200
Accounts receivable	30,000	32,500
Inventory	40,000	42,300
Fixed assets	25,000	31,000
	$100,000	$110,000
Liabilities		
Accounts payable	$ 26,500	$ 34,500
Mortgage payable	10,000	10,000
Capital stock	50,000	50,000
Surplus	13,500	15,500
	$100,000	$110,000

From this we see that the net worth of the business has increased by $2,000, but the balance sheet does not show us what brought this about. Nor does it tell us that $3,000 was earned, declared, and paid out in the form of dividends. For this and other facts, we must have recourse to the income account. Neither the balance sheet nor the income account by itself is sufficient; together they show us in what respects the business has changed and what has caused these changes.

Accounts of individual proprietorship and partnership. In the foregoing analysis of business accounting, we have used examples of the balance sheet and profit and loss statement of a corporation. The same accounting principles and practices apply also to the individual proprietorship and the partnership. As regards the balance sheet, the only difference is that the statement of the individual proprietorship or the partnership would not contain the item, capital stock, or surplus. Instead there would ordinarily be simply the term "capital." The income account of the individual proprietorship or the partnership is in form identical with that of the corporation.

EXERCISES

1. Arrange the following items in the form of a statement of the assets and liabilities of the Bolter and Hammer Hardware Co., as of December 31, 1946: land and buildings, $90,000; capital stock, $100,000; 5 per cent mortgage bonds outstanding, $50,000; machinery and equipment, $75,000; government bonds, $5,000; cash, $7,000; surplus, $66,740; taxes due, $3,260; accounts receivable, $12,000; raw materials on hand, $26,000; finished goods on hand, $20,000; accounts payable, $15,000.
2. There are 1,000 shares of capital stock outstanding. What is the par value of each share? The book value?
 Note: This and each of the following five exercises is to be regarded as an independent, not successive, transaction. In each case, refer to the original statement in Exercise 1.
3. A cash dividend of 6 per cent is declared and paid to the stockholders. Make the necessary changes in the statement. What is the book value per share of stock now?
4. The company declares a stock dividend of 50 per cent. Make the necessary changes in the statement. What is the new book value per share? Stockholder A held ten shares of capital stock before the stock dividend was declared. Calculate the book value of his holdings of capital stock before and after the stock dividend.
5. A fire breaks out and destroys the stock of raw materials and the warehouse, valued at $5,000, in which they were stored. These losses are not covered by insurance. Make the necessary changes in the statement. Is the company still solvent?

6. Make such changes in the balance sheet as would be necessary to illustrate stock-watering.
7. The general ledger of the Bolter and Hammer Hardware Co. contains the following income and expense accounts:

Cost of goods sold	$32,650.00
Advertising	250.00
Salaries of office force	1,240.00
Workshop income	110.00
Interest paid	125.00
Sales (net)	47,375.00
Wages of sales force	4,650.00
Office supplies and postage	70.00
Delivery expense	950.00
Interest received	40.00
Rent paid	1,200.00
Miscellaneous expenses (general)	500.00
Depreciation of store building	130.00
Dividends declared	2,500.00
Taxes	559.50

From the above facts prepare a profit and loss statement for the company for the year.

CHAPTER IV

Large-Scale Production and Combination

The rise of the standard of living. It is a fact familiar to every reader of history that in the past century and a half there has been an increase in the material prosperity of this and other industrialized countries that is without parallel in the records of mankind. The factory system, with the application of power to machinery, resulted in an outpouring of an amazing flood of goods; articles that had been luxuries destined for the consumption of the few began to appear even to the poorest families as necessities, while thousands of new goods—both capital goods designed to increase the efficiency of labor, and consumers' goods to fill wants unknown to our forefathers—came forth in a steadily increasing stream.

This growth in national income has brought to millions of families improvement in the scale of living and release from excessive hours of toil. Even as recently as the middle of the last century, factory laborers in the United States worked an average of 69 hours a week—almost 12 hours a day for 6 days a week. Gradually as the average product of the workers increased, hours of labor decreased; by 1920 they were down to 48.2 hours per week, and by 1940 to 39.4 hours.[1]

Accompanying this decrease in hours of labor, which for the period from 1850 to 1940 amounted to 43 per cent, has been an increase in the return to labor, both in terms of money wages and in what money would buy. A striking illustration of this is seen in the rubber tire industry. In 1908, when a tire with a life of about 2,000 miles cost $35.00, or 1¾ cents a mile, the average wage in a certain tire factory was 40 cents an hour. One hour of labor would therefore pay for only 23 miles of use of that one tire. By 1936 the hourly wage had gone up to 88 cents, an increase of 120 per cent in the money wage rate, but the price of the tire had dropped to $8.00 and its life had increased to an average of 20,000 miles. One hour

[1] W. I. King, *Raising the Workingman's Scale of Living.* New York: Committee for Constitutional Government, 1945, p. 7.

of labor would in 1936 pay for 2,200 miles of use—an increase of ninety-five fold. This in part explains why the workingman could not own a car in 1908 and why he could in 1936.

For the period from 1850 to 1940, the index number of real wages rose from 100 to 446, and in the light of this fourfold increase it is quite understandable that a growing share of the income of the laborer could become available for other things than the routine needs of life—for education, medical care, travel, amusements, and durable consumers' goods such as electrical appliances, automobiles, and the like. According to a report of the National Industrial Conference Board, "about a third of the wage earner's outlay is now allocated to such items, as compared with a little more than a tenth in 1869–1879." [1]

The growth of capital. Blessed with ample natural resources and a virile and adventurous people accustomed from colonial days to independence of thought and action, with a tradition of political and economic freedom firmly established, a population increasing because of a high birth rate and immigration from abroad, and nearly half a continent without artificial trade barriers, the United States provided an environment peculiarly favorable to industrial development.

It is, however, well to remember that the large industrial plants of today, which we accept as a matter of course, had very humble origins. In the early nineteenth century, factories were small and machinery was neither plentiful, efficient, nor reliable. Savings available for transformation into capital goods were not abundant; nor despite the brilliance of an Eli Whitney, had the American inventive genius really manifested itself. The beginnings were small; and progress at the start was slow and conditioned on the gradual accumulation of capital, on the progress of invention, on the growth in the size and organization of the producing unit, and on the development of improved transportation and distributive systems.

Figures of the growth in the volume of capital per worker show more clearly than would figures of total capital investment the increasingly "capitalistic" character of industry. The chart on page 58 gives these for manufacturing in the United States during the period from 1849 to 1942. Starting with a figure of about $600 in

[1] National Industrial Conference Board, Inc., *Productivity and Progress.* New York: the Board, 1946, p. 26.

1849, there is a more or less steady growth in the amount of invested capital per worker until a peak is reached in the early thirties of something over $8,000, with a subsequent decline to $5,800 in 1942. As the worker became more liberally supplied with capital, there was increase both in the output of goods attributable to him and in his wages. On the chart will also be found curves showing both the growth in the value added to the goods in

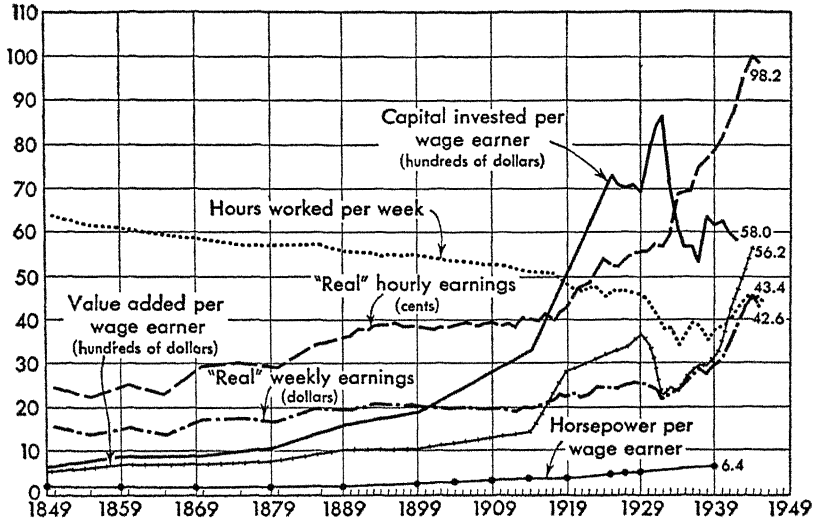


FIG. 1. INDICES OF AMERICAN MANUFACTURING, 1849–1942 [1]

the process of fabrication and the increase in real wages, it is interesting to note the fidelity with which in general these follow the curve of invested capital. That this relationship between real wages and capital investment is not mere coincidence is shown by comparable figures for other countries. It is estimated for example that in 1930 the horsepower per worker in Great Britain was 52.7 per cent of the figure for this country and in Germany 53.7 per cent; other figures would show both total capital investment per worker and real wages to be materially below those of the United States.[2]

[1] Figures furnished by the National Industrial Conference Board.

[2] Carl Snyder, *Capitalism the Creator*. New York: The Macmillan Company, 1940, p. 259.

Capital funds for investment in business come either from the savings of individuals or from the undistributed profits of corporations. Historically, the savings of individuals have been of paramount importance, but in recent years the reinvestment of undistributed corporate profits has become of increasing significance.

The wealthier individuals have always provided the largest part of the funds for investment, for as a general rule the larger the income the greater the proportion saved; but that the relatively small investor is not a negligible figure is shown by the fact that, in 1941, 34 per cent of all dividend receipts reported to the Bureau of Internal Revenue by individuals or fiduciaries went to those who reported taxable net incomes of $5,000 or less.[1]

Large-scale production: Importance of large plants. A large plant is not an essential part of the capitalist method of production; for there are many small plants today employing relatively small labor forces, yet liberally supplied with tools, machinery, and power. Nevertheless, the growth in the average size of the producing unit and the development of the very large producing unit may properly be considered one of the main factors underlying our rapid industrial advance and growth in national income; for so many advantages may be derived from concentrating a large number of workers and a large amount of equipment in one establishment under unified management that in many lines of industry the compulsion to expand the size of the plant has been irresistible.

Taking the cotton manufacturing industry of the United States as an example, we find that between 1859 and 1925 there was an increase of 55 per cent in the number of cotton manufacturing plants, the average number of workers per plant had more than doubled, the average primary horsepower per establishment (an index of the use of power-driven machinery) had increased enormously, and the average value added to the product per establishment had increased by 334 per cent. If we were to examine the statistics showing the growth of other industries, we would find differences in detail but the same general trend—a great increase in the average number of workers, the average amount of capital, and the average output per establishment.

The very large establishment has attained great importance in the present-day industrial life of the United States. According to

[1] Committee on Postwar Tax Policy, *A Tax Program for a Solvent America*, New York, 1945, p. 71.

the figures of the 1939 census, industrial establishments with an annual product valued at $1,000,000 or over formed only 5.2 per cent of the total number of such establishments, but they were responsible for 67.5 per cent of the total value of the product and employed 55 per cent of the total wage earners in industry. On the other hand, establishments with an annual output of less than $100,000, while actually forming 69.5 per cent of the total number of producing units in the year 1939, accounted for only 6.8 per cent of total industrial production and employed but 11.5 per cent of industrial labor.

In 1939 the establishments employing from one to five laborers comprised 41.2 per cent of the total number, whereas the establishments employing over 500 workers formed 1.2 per cent of the total, but these large establishments employed 35.4 per cent of the total wage earners in industrial plants, as contrasted with the 2.6 per cent employed by the establishments in the former group. Establishments employing more than 1,000 laborers formed only 0.44 per cent of the total number of establishments, but they employed 22.4 per cent of the total number of workers.

Two conclusions of considerable importance can be drawn from these figures. The first is that small-scale industry is still of significance in our national life, particularly to supply local markets. Big business has not entirely supplanted the small producer, even in the lines of industry where the advantages of large-scale production seem most marked. At the same time, it is true that the large units have been growing, both absolutely and relatively, at the expense of the smaller units. Production for national and international markets has been supplanting production for local or regional markets, and with this change has come the decline of the small unit and the substitution of the larger.

Conditions necessary for large-scale production. The development of large-scale production depended first of all on growth in technical knowledge and on the progress of invention, for in many lines of industry it could not be introduced profitably until the machine-tool industry had advanced to the point where it could design and build machines capable of producing standardized and interchangeable parts.

In the United States pioneering in making interchangeable parts is found, even before the middle of the last century, in the manufacture of firearms. Shortly thereafter the principle was adapted to

the manufacture of other goods, notably watches, sewing machines, agricultural machinery, and bicycles; and by the turn of the century it had found general application. At the same time the invention of machinery of increasing power, capacity, precision, and versatility added strength to the trend toward the establishment of larger producing units.

A second prerequisite for large-scale production is obviously a product that is in great demand, or one for which a large demand can be created; for unless a large output can be marketed profitably, there is little scope for increased refinements in the productive apparatus. The growth in the population of the United States in the past century, together with the increasing purchasing power of the individual, provided ideal conditions for the disposal of an increasing volume of goods.

A third requirement is that the article be not one which is demanded because of the finish which hand labor alone can give it. It must be amenable to the machine process, capable of standardization, and attractive in the eyes of the purchaser despite standardization. Many of the final steps in the manufacture of articles of luxury and of personal adornment, such as the finer grades of clothing, can never be handled by large-scale methods.

Manufacturing is the field in which large-scale production obviously finds its greatest opportunities and from which may be drawn the most illuminating illustrations of its principles. Many other lines of industry have, however, proved themselves amenable to large-scale production.

In transportation, banking, and marketing the size of the organization has greatly increased in the last century. Doubtless this increase in size has come partly in response to growth in the size of the manufacturing establishments and is partly a cause of that growth. In any event concentration of manufacturing in large plants requires more effective service than could be afforded by the transportation, banking, or marketing system of a century ago.

The one great industry which has not proved a profitable field for the application of large-scale methods is agriculture. There are of course farms where hundreds of workers are employed, but they are the exception and not the rule.

The economies of large-scale production: Management. The adoption of large-scale methods has not come about automatically, but because some entrepreneur has seen in the enlargement of the

producing unit possibilities of reducing unit cost and of increasing profits. Furthermore, the economies of large-scale production are not assured, even though a large plant be established; the opportunity of securing economies exists, but whether potential economies are transformed into actual savings depends on the intelligence, judgment, imagination, and ingenuity of the managerial staff. For these reasons it seems appropriate to assign to management the first place among the sources from which such economies may arise.

There are some areas in a plant where the administration and supervision are pretty much a matter of established routine. Here savings may flow almost automatically to a large plant; for while total costs of administration will be larger than for a small plant, they will not be proportionately larger. Within limits the heads of some departments can handle a large department about as efficiently as a small one. The chief additional expense will be for an enlarged clerical staff.

But if a factory is to keep ahead of its competitors and to make constant progress, its chief executives—the ones who do the planning and determine major policies—must possess managerial ability of a high order. Men of outstanding capacity seem not to exist in sufficient numbers to fill the key positions in all plants; and because the large plant can spread costs over a greater volume of goods, it can afford to offer the very ablest men salaries that would be prohibitive for most small plants.

Where a small factory may assign several functions to one man, the larger plant may have one man or several men specializing in one field, and the advantages that flow from this specialization may be further heightened by the stimulating influence that each man has on the others. The aggregate abilities of a well-selected staff should be measurably greater than the sum of the individual abilities.

The cost of buying raw materials may be less per unit for the large factory than for the small one. Buying in large quantities means a real saving to the seller in packing, shipping, and sales cost, and he can therefore afford to quote a lower price to one who buys on a large scale. If the head of the purchasing department is dealing with one source of supply, he can, by placing advance orders, coördinate the production schedule of the supplier with that of his own concern, with profit to each.

Similarly, in the sale of the finished products, economies are also possible. Selling in a national or an international market may make possible better coördination of the activities of salesmen, thus covering a given territory without overlapping and with greater effectiveness; direct sales organizations may perhaps be established abroad, and national advertising can be used with greater results.

Power and machinery. One great economy of large-scale production results from saving in the use of power-driven machinery, both in the use of power as well as in the use of machinery. The saving in power arises from the lower unit cost of installing and operating a power plant. Though a large plant costs more to install than a small one, it costs proportionately less per unit of output, and greater efficiency in the production of power results. Hence the cost of power to be attributed to each finished article is less.

Also machinery can be used for more of the processes in a large factory than in most small factories. A machine, to be used to advantage, must be used more or less continuously. If a machine is used only a part of the time, the cost which must be borne by each unit of the product undergoing the particular process may make the use of the machine more expensive than hand labor or some alternative method.

It should be emphasized that modern heavy machinery, with its capacity for performing with amazing accuracy and speed several simultaneous operations, is very expensive—one need but walk into an automobile plant or a steel mill to appreciate this—and that not infrequently such machines require the use of expensive accessories. It is stated that the annual cost to a manufacturer of the dies, jigs, and fixtures used in shaping automobile bodies may vary from $2,500,000 to $6,000,000—a prohibitive cost unless the plant has a large output.

There may be further savings in the installation, care, and operation of a number of machines in batteries, if the scale of production is sufficient to warrant it. This seems to be particularly true of automatic or semi-automatic machinery, which allows one man to tend several machines at the same time.

Technical experiment and research. A large company is better able to conduct experiments than a small company. Experimentation is a risky and expensive affair. Many trials are necessary before a device or a process is developed to the point where it is of practical use. For a small concern the financial burden may be

too heavy to carry, whereas the company with greater resources can carry the load through the lean years, trusting that in the long run the improvement will more than pay for itself.

By-products. Another economy lies in the utilization of by-products. The meat-packing industry offers one of the most illuminating examples of this. According to figures given by Swift and Company, by-products such as hides, fats, etc. form 45.7 per cent of the total weight of a live steer. A list of the finished goods which are made in whole or in part by Swift and Company from these by-products includes the following: glue, fertilizer, oleomargarine, lard compounds, soaps, glycerin, leather dressing, animal feed, buttons, combs, gelatin, ice cream, candy, pharmaceutical products, snuff containers, goldbeater's skins, and sausage containers. In the early days the materials from which these goods are now made had to be carted from the city and buried; so far from being a gain, they were the occasion of considerable expense.

The economy which results is plain; any net returns obtained from the by-products diminish by so much the price which must be obtained for the principal product in order to cover the expenses of production. The meat packers, for example, often pay more for live cattle and hogs than they receive for the dressed meat, but they make up the difference in the return from by-products.

The place of small-scale production. The advantages offered by large-scale production seem so great that a question often arises as to how the small plant manages to survive (as it does in virtually every major branch of industry) alongside the large plants making the same product. For example, in steel works and rolling mills there were, in 1937, 67 establishments, 16.3 per cent of the total, which had an average working force of 100 or less; and even in the manufacture of automobiles, in which the very large plant is so predominant, there were 46 establishments, 35.1 per cent of the total, with an average number of wage earners of 100 or less.

In some cases the answer is undoubtedly found in production for a local market. Or it may be that the plant has unusually favorable access to the raw materials. In other cases it is possible that the adoption by the larger plants of a policy of "live and let live" has kept the prices of the products higher than necessary to meet the costs of the large plants and still high enough to permit the more efficient of the small plants to survive. But these explanations do not entirely cover the situation.

Just as there are large plants that do not reap the full economies of large-scale production because they are making too many different items, so there are small firms that have secured these economies by intense specialization. Concentrating exclusively on the manufacture of one article or of one part, they can provide continuous operation for their machines and specialization for their workmen. They may even be able to avoid many of the normal sales costs by producing entirely under contract for a mail-order house or for another manufacturer. It is said that in the automobile industry many of the companies with less than 100 employees each are realizing in this way all of the economies of large-scale production.

It is also true that weakness or lethargy in management provides the answer in many cases. In pointing out the advantages the large plant may have, efficient management was given first place. But it does not necessarily follow that the large plant will always secure the best managerial staff; in some cases it obviously does not. Many factors other than the possession of superior business ability may enter into the choice of the principal officers of a plant. And it is certainly far from the truth that inferior managerial ability is characteristic of all small industry. The fact that profitable small concerns earn a higher rate of return on invested capital than the larger companies is sufficient to dispel any such notion.[1]

Importance of the large corporation. We have become so used to the role played by the great corporations in our everyday life that, paradoxically perhaps, it is sometimes rather surprising to discover how large these corporations loom in the economic picture.[2] According to one estimate the two hundred largest nonbanking corporations in the United States had at the beginning of 1930 total assets amounting to $81,074,000,000; the assets of all the nonbanking corporations, something over 300,000 in number, amounted to $165,000,000,000. Almost half of the assets of all nonbanking corporations were in the control of less than 1 per cent of the total number of such corporations. The same authorities also estimate that these corporations owned 38 per cent of all business

[1] Crum, W. L., *Corporate Size and Earning Power,* Cambridge: Harvard University Press, 1939, p. 28.

[2] The figures given in this section on large-scale production cannot be compared with figures in this section since the former are given in terms of "establishments"; an establishment is defined as a single plant or factory, though in some few cases it may include more than one plant.

wealth, other than banking, and 22 per cent of all national wealth.[1]

A somewhat different picture of the importance of large corporations is given in the following table, which shows the number of corporations in selected lines having in 1929 an income of $1,000,000 or more and the percentage of total net income in the given line which accrued to these large corporations.[2]

CONCENTRATION OF CORPORATE CONTROL, 1929

Industry	Number of corporations with net income of $1,000,000 up	Net income (*In millions*)	Per cent of all net income
Manufactures	627	$3,338	64.0
Mining	65	278	84.6
Public utilities	230	1,805	86.0
Trade	93	316	27.5
Service	31	108	34.4
Finance	283	1,048	47.7
Total	1,329	6,893	60.5

In recent years the large corporations have shown a growth which is much more rapid than that of the small or medium-sized corporations. From 1907 to 1927 the assets of the 200 largest nonfinancial corporations increased more than twice as rapidly as the assets of other nonfinancial corporations; they reinvested a larger portion of their earnings; they secured a larger share of new capital by the issue of securities; and mergers were more common and more significant among them.

Large as well as small corporations suffered severe losses during the depression years of the 1930's, but the large corporations emerged at the end of the decade in stronger financial condition than did the small corporations. During the so-called defense and war years from 1940 to 1945, the exigencies of the national production program undoubtedly favored the larger producers and gave rise to widespread concern over the fate of small business. Even during the years of depression, mergers were being consummated, although at a slower rate than occurred before 1929. The Federal Trade Commission reported in 1947 that the war accelerated the ac-

[1] For a further discussion of this topic and for an illuminating analysis of the social implications of the position of the modern corporation, the student is referred to Berle, A. A., and Means, C. C., *The Modern Corporation and Private Property,* New York: The Macmillan Company, 1933, from which these figures are taken.

[2] Corey, Lewis, *The Decline of American Capitalism.* New York: Covici, Friede, Inc., 1934, p. 387.

quisition of small firms by large corporations.[1] Larger manufacturing concerns were reported to have absorbed over 1,800 formerly independent concerns, the assets of which amounted to $4,100,000,000, or about 5 per cent of the assets of all corporations in 1943. The rate at which mergers have taken place since the end of the Second World War is said to be the highest since 1929.

Combination: Definition. Rarely does large-scale production give rise to problems of vital concern to society or raise the question of regulation in the public interest. Sentimentalists deplore the elimination of many small shops and the passing of the independent craftsman. But while we may have lost thereby a distinctive and virile element in our economic life, few would seriously suggest that we sacrifice the greater gains which have come from mass production in order to try to recapture the spirit of an earlier day.

This, however, is not always true of the *combination,* which we may define as *a union of several industrial units under a common management.* There are indeed serious problems of contemporary interest and importance raised by the great corporations, the combinations, and the combinations of combinations.

The formation of combinations. Many a corporation has become a combination quite simply and naturally when it was found profitable to build a new plant from earnings or from funds secured by the floating of additional securities. Some have grown by the consolidation of two or more existing organizations; one company has purchased the property of another company and then completely merged the two into a larger single company. The American Tobacco Company, formed in 1890, exchanged its stock for the plants, business, brands, and good will of five companies, and the American Sugar Refining Company, organized in 1891, is another example among the many that might be cited.

The most common, the most effective, and the least expensive method of building up a large combination is, however, through the holding company. A corporation which buys the controlling shares of stock in another company, either with its own stock or with cash, becomes in effect a holding company. Frequently a new

[1] Federal Trade Commission, *The Present Trend in Corporate Mergers and Acquisitions,* Washington, 1947. For other recent discussions of this problem, see National Resources Committee, *The Structure of the American Economy,* Washington, 1939, and Temporary National Economic Committee, *The Distribution of Ownership in 200 Largest Non-financial Corporations,* Monograph No. 29, Washington, 1940.

corporation is chartered, as in the case of the United States Steel Corporation, to buy up a controlling number of shares in the companies which are to be combined. Little capital is required to start a holding company, since the shares of stock of the member companies may usually be obtained in exchange for the stock or bonds of the holding company. Unity of policy may be obtained by the election of the proper men as directors in the constituent companies.

Motives leading to the formation of combinations. In considerable measure combination has been resorted to as a means of securing substantial control of a particular field of production in order to obtain monopoly profits. Two other motives, however, have been prominent. Of these the first has been to make possible speculative profits through the issuance of large volumes of stock; the second has been to obtain even greater economies in production than were possible simply through large-scale production.

We have, therefore, three distinct problems. The first, and in some aspects the most important, concerns itself with the growth of monopolies—the so-called "trusts"—and their effect on society and the methods by which they may be regulated in the public interest. The second involves methods used in marketing securities and the means by which the operations of investment bankers, underwriters, and the security exchanges may be controlled in order to prevent abuses.

The third problem, to which the remainder of this chapter will be devoted, centers around the possible social advantages of combinations as producing units. Are there economies in production possible to a number of large plants under single management which cannot be obtained by a single large producing unit? If there are such economies and if they are of considerable significance, this fact will have an important bearing on the monopoly problem.

Horizontal combinations: Definition. We may distinguish two types of combination: *horizontal* and *vertical*. As the name suggests, a horizontal combination is one which unites under a single management several plants producing the same product or products. It is a type quite common in the United States, being exemplified by many of the trusts which exist or have existed in the past, such as the American Sugar Refining Company, the United Shoe Machinery Company, the American Tobacco Company, and others.

Economies in operation. The greatest field for economies would appear to lie in the operation of the plants. Each plant can presumably secure maximum efficiency by producing a certain number of articles. If forced above that point or below it, the unit cost of goods is increased. With several plants operating under unified control, it is possible so to arrange things that all the slack in production will fall on one or more plants, leaving the others running as usual. The American Sugar Refining Company is said to have used its Brooklyn refinery for this purpose, slowing down operations there as demand slackened and speeding up in response to increased demand.

In some industries the risks and uncertainties involved in production may be reduced if plants are located in different parts of the country. A labor strike may not affect all of the plants at the same time, and a tie-up of the transportation facilities of all the plants at one time is unlikely, so that if the closing of some is forced the others can be operated more intensively.

Other operating economies are available to a combination of plants that cannot be readily effected in a single plant. A combination which makes a variety even of the same general product, such as tin cans, is able to specialize its factories, devoting one factory to the making of a particular article, rather than allowing all factories to make all articles. In this way the utmost advantages of specialization are attained. The devices or patented articles that have proved most successful can be made the common property of all the operating units, heightening the efficiency of each. With several plants it is possible to start competition between them, thus enlisting the enthusiasm of one manager to excel the others. Cost statistics may be available for all the plants, so that if one is unsuccessful in some respect the trouble can quickly be diagnosed and remedied. Interplant competition may disclose men who are most capable of handling the more difficult tasks of coördination and centralization.

Economies in selling. Similarly, the large corporation enjoys certain advantages in the sale of its products. Greater volume and lower selling costs per unit of sales may result from being able to offer customers a full line of similar goods.

More effective use of salesmen and other sales media is likely to result as new products are added to the line carried. If the manufacturer is selling his product under a trade name—and where

feasible this is the practice—the unit cost of making the public conscious of the trade name, when spread over a variety of household electrical appliances, let us say (such as stoves, vacuum cleaners, toasters, irons, refrigerators, fans, and radios), may be much less than if the effort were limited to electric stoves alone. Too great a diversity in production may of course make it impossible to secure the full advantages of large-scale production, though some concerns seem to have been able to achieve economies, both in production and in marketing, by intensive cultivation of both the domestic and the foreign markets.

An illustration of economies in selling may be seen in the experience of the tobacco trust. This had been formed by the combination of several large producing units and was dissolved by a court decree in 1911, whereupon the business was turned over to a number of separate companies. "The subdivision of the business provided for in the dissolution decree led to a duplication of selling organization and an increased overhead expense; and the result was a general increase in selling costs. . . . In every branch, except flat plug and Turkish cigarettes, the selling costs showed an increase. . . . The advertising expenditures of the successor companies also greatly increased as compared with the expenses of the trust. The advertising expenses of the trust in 1910 in all branches except cigars were $10,895,132, while those of the successor companies in 1913 amounted to $23,623,564, or more than double." [1]

Economies in buying. A horizontal combination may possibly secure a saving in the purchase of its raw materials. Buying in much larger quantities than the single plant, it may approach the producer directly and realize for itself the middleman's profits, or it may secure its goods from a wider productive area and thus obtain its raw materials more cheaply. But it is to be doubted if these economies are great enough to be of real consequence.

Savings in freight costs. The savings on cross freights are sometimes cited as proving the superiority of the combination. If under separate management there are two plants, one in New York and the other in Chicago, each may sell goods in the other's district and incur freight charges unnecessarily great. Under combination the goods will be sent from the nearest point. But the importance of this saving, while real in many cases, can be greatly exaggerated;

[1] Jones, Eliot, *The Trust Problem in the United States*. New York: The Macmillan Company, 1921. pp. 149–50.

for if the goods be bulky and of low value the distance they can be shipped is limited by the cost of freight, whereas if they be small and expensive objects such as watches, freight charges form such a small portion of their value as to be of little moment.

Vertical combinations: Definition. Vertical combination involves the gathering under one management of the various processes necessary to turn out a finished good. It is sometimes called *integration* of industry and is the opposite of that process of differentiation and specialization which has been characteristic of the factory system.

In some industries there has always been a certain amount of integration. Thus in the United States spinning and weaving have commonly been undertaken by the same factory, although in England they are separate industries. In the steel industry of the United States can be found the most impressive illustrations of vertical combinations, although horizontal combinations also have been common.

Technical advantages. There are real technical advantages in combining thus the different stages of manufacture of steel products, advantages which result in saving of fuel for reheating the metal, saving of time and labor in moving or handling material, and more complete utilization of by-products, such as blast-furnace gas. Further advantages may be obtained through the control of sources of iron and coal.

The steel industry differs from some others in that each unit not only may use the products of another as raw material but may pass its own finished or semifinished products on to serve as raw materials for the others. Coal and iron are essential for the blast furnaces, while the ultimate products of the blast furnaces are requisite in the mining and transportation of coal and iron.

The United States Steel Corporation. The extent of the process of integration in the United States can best be pictured by a brief summary of the organization and scope of operations of the United States Steel Corporation, one of the largest corporations in the United States, incorporated in 1901 with an authorized capital stock of $1,100,000,000. The corporation secured control of a number of other corporations by an exchange of stocks and bonds, and at the outset it controlled eleven corporations. It is a fact of interest that almost all the companies which were united to form the United States Steel Corporation were themselves combinations. Some of

them represented merely the horizontal type of combination, such as the American Steel Hoop Company, which was a consolidation of nine concerns formed with the desire to limit competition and secure larger profits; while some already showed signs of a fairly complete integration, as the Federal Steel Company and the Carnegie Company.

The United States Steel Company was not exclusively a vertical combination, for several of the constituent companies produced the same articles. The result of the consolidation was the creation of a giant corporation, which was practically self-sufficient so far as materials were concerned, and which embraced the manufacture of practically every known article in steel and iron.

Conclusions. While there is little doubt that both vertical and horizontal combination offer possibilities of economies over and above those that can be obtained by the single large manufacturing plant, it is exceedingly difficult to determine the extent to which these economies have in actual fact been realized, for in few cases are comparative figures of costs of production available. A high rate of return on invested capital cannot be accepted as completely valid evidence of economical operation, for if a large combination were to have conspicuous financial success, it might be possible that the record of profits could more fairly be attributed to monopolistic control over production and high prices than to economies in production.

Nevertheless, were it true that the largest combinations invariably or usually showed rates of return in excess of those previously earned by the member companies before merger or of large single plants in the same line, it would, in the absence of proof to the contrary, be natural to give some credit for the increased rate of profits to economies of combination.

As a matter of fact, however, the success of the large combinations in the United States has not been striking, if success be measured in terms of profits. Not only have profits been smaller than anticipated by the promoters of the consolidations, but in many instances they have been smaller than the aggregate profits of the constituent companies in the period prior to consolidation.

For example, the United States Steel Corporation during its first year of operation earned but 78 per cent of the amount earned by the member companies in the period immediately preceding the formation of the combination, and for the first ten years of its life

its average earnings were somewhat less than its earnings during the first year. Professor Dewing has made a study of the earnings of a number of mergers and found that the earnings for the first year, as well as the average earnings for the first ten years of the existence of the combinations, were materially below the level of the previous earnings of the member companies.[1]

A more recent study made under the auspices of the Temporary National Economic Committee as to the relative efficiency of large, medium-sized, and small business confirms these findings In summarizing the results of the investigation the report states that "in the 233 combined tests, large size, whether represented by a corporation, a plant, a group of corporations, or a group of plants, showed the lowest cost or the highest rate of return on invested capital in only 25 tests. In these combined tests, medium size made the best showing in 128 tests and small size in 80 tests." [2]

In his study of the profitableness of corporate business Professor Crum found that "on the average large enterprise . . . is more profitable than small enterprise, especially very small enterprise." On the other hand, when only corporations that showed a profit were analyzed, the highest rate of return was earned by corporations in the smallest size class, and in all years covered by the study save one, the minimum rate appeared for the largest size class.[3]

The failure of most large combinations to give evidence of economies of operation and management in their earnings indicates that fundamental weaknesses are a consequence of size of organization. A huge combination is vulnerable in many ways, but without question the most difficult problem it has to face is that of finding men of the caliber required to direct the work. When one reflects that, in theory at least, the directors and officers are expected to be familiar with the problems of many widely scattered plants producing a variety of goods, one wonders whether there exist human beings competent to undertake the responsibilities involved and carry them out effectively At any rate, business ability of the very highest order is essential if economies from combination are to be secured, and the evidence does not show that such ability is suffi-

[1] Dewing, A. S., *The Financial Policy of Corporations*. New York: The Ronald Press Company, 1942, Fourth ed., Vol. II, p 931.

[2] Temporary National Economic Committee, *Relative Efficiency of Large, Medium-sized, and Small Business*, Monograph No. 13. Washington, 1941, p. 14.

[3] Crum, W. L., *Corporate Size and Earning Power*. Cambridge, Harvard University Press, 1939, pp. 7, 28, 230.

ciently plentiful to assure that large combinations will, by more efficient production, be able to undersell the large, efficient, single units.

This conclusion does not give warrant for the inference that the day of the large combination is over. On the contrary, many indications point to a continued process of growth by merger, so long as no legal obstacle interposes itself. The acquisition of monopoly power or the elimination of intense competition or the opportunity to secure profits from the sale of corporate securities may provide sufficient motives for management to try to expand the scope of corporate activities, whether or not the results be consonant with the interests of society as a whole.

CHAPTER V

Economic Freedom and the Control of Production

The control of production: Conscious and unconscious control. When we survey production as a whole we see the truly marvellous spectacle of millions of men engaged in thousands of different tasks, turning out quantities of goods varying in kind from the most essential to the most frivolous, transporting these goods where they are most urgently wanted, financing their manufacture, shipment, and sale, and performing personal services of the most varied character. Each producer is apparently quite independent in his planning, yet each attempts to turn out exactly the amount of his particular product which can be sold, and with few exceptions the goods all seem to reach the consumers where and when they are wanted.

This extraordinary coördination of human activity cannot be the consequence of accident, nor is it merely a fortunate coincidence; there must be some control. Now such control may be of two types. The first type we find exemplified today in Soviet Russia, where legal authority restricts the freedom of the individual producer and prescribes in minute detail the character and scope of individual activity. The second type of control is exercised by economic forces operating under a régime of personal liberty. We may call it unconscious regulation in contrast to the first, which is consciously exercised.

Mercantilism. This second form of control represents a reaction against the restrictions of an earlier system of conscious control. In this earlier period the economic and political philosophy, to which the name *mercantilism* is given, held that regulation of the activity of the individual was a necessary function of the state, partly in order to protect the subjects of the state against the dishonesty or oppressive power of their fellow citizens or of foreigners, and in part to increase the prosperity and prestige of the nation.

In harmony with the first of these purposes we find, from an early time well into the eighteenth or even the nineteenth century, laws

and regulations restricting the freedom of action of the individual. In England there were, for example, laws fixing the price of bread and the rates of wages, laws providing for the inspection of cloth, laws regulating the conditions of apprenticeship, laws prohibiting the export of bullion and of certain raw materials, laws restricting the movement of labor from one part of the country to another; while in France during the period of most minute regulation the individual could not determine what he should produce, how much he should produce, where he might produce it, or to whom he might sell it.

The second motive leading to regulation was the desire to increase the prosperity and power of the state. During the sixteenth and seventeenth centuries the nationalistic spirit was manifesting itself to an ever-increasing degree, and it was generally agreed that the first duty of the ruler was to expand his power even at the expense of his own subjects as well as of foreign nations.

The political and economic philosophy of mercantilism was a part of the program of promoting the national welfare. Special attention was given to the regulation of the nation's foreign trade for the purpose of strengthening the nation for armed conflict with rival powers. Thus gold and silver were desired, and attempts at prohibition of the export of the precious metals were made. These efforts generally resulted in failure. Attention was then directed to regulating the country's imports and exports of merchandise so that a "favorable" balance of trade would be attained and gold and silver would be drawn in from other nations in payment for goods.

***Laissez faire:* Reaction from government control.** During the eighteenth century the conviction spread that there had been too much government, too much interference with private business, and there arose a strong reaction against the restrictions of mercantilism.

Laissez faire—let things alone—became the watchcry of philosophers, economists, and statesmen. *Laissez faire* in the first place emphasized the importance of the individual and of individual welfare. Secondly it assumed that individual welfare and national prosperity are not incompatible.

It went further and asserted that individual prosperity is at the very foundation of national prosperity; that the individual, if allowed to follow his own self-interest without external restriction,

would exert himself to the utmost in the line of endeavor for which he was best fitted and would therefore be a more productive and hence a more valuable member of society than if his occupation and his conduct in that occupation were prescribed for him by either law or custom. As each man followed his own interest, the clash of conflicting interests would lead to modifications and compromises which would result in approximate justice to all.

Translated into terms of concrete policy, *laissez faire* supported (1) free trade, as opposed to protective tariff, (2) no attempt to fix or regulate prices or to regulate either the quantity or quality of goods produced, (3) freedom on the part of the individual to enter any occupation he might choose, in any place he might think advisable, (4) the determination of wages by unfettered bargaining between employer and employee, (5) the limitation of the functions of government to (*a*) defense and the maintenance of peace and order, (*b*) the dispensing of justice, (*c*) the establishment and maintenance of certain public works and institutions, such as the paving of streets, the improvement of harbors and means of communication, the care of education, etc., which could not be made commercially profitable as private enterprises.

The reaction against mercantilism spread rapidly in the eighteenth and nineteenth centuries, and the general acceptance, particularly in England, of the doctrines of *laissez faire* ushered in a period in which the dominant note was the endeavor, so far as possible, to give free play to the forces of individual initiative. Despite the fact that the scope of government regulation has at many points been broadened beyond the limits of extreme *laissez faire,* the conditions under which production has operated have on the whole been those of economic freedom in contrast to legal authority. Taking a broad view of the system which developed under the influence of the doctrine of *laissez faire,* we note that it is founded on three major principles: (1) personal liberty, (2) private property, and (3) individual initiative and control of enterprise.

Personal liberty. So generally is personal liberty taken for granted in America today that we are in some danger of failure to appreciate its full significance. In truth personal liberty is a comparatively new thing, only recently gained as the climax of a struggle of centuries. Throughout most of the world's history freedom was a rare privilege enjoyed by the favored few. The masses of the people were slaves or serfs, more or less completely subject to

the will of their owners or overlords. Even after slavery was abolished and the status of serfdom gradually liberalized and finally developed into comparative personal freedom, the legal restrictions upon economic activity which we have been studying continued. Real individual freedom is scarcely a century old.

Private property. The institution of private property also is a remarkable and fairly modern institution. There was probably very little private property in ancient days, common ownership by the tribe or nation or ownership by the chief or king being the rule. The feudal system was based upon ownership of at least the most important element of the wealth—the land—by the king or other ruler. Even today a certain part of the wealth of every community, more in some countries than in others, is owned in common as the property of government. But in the United States the great bulk of all the wealth is owned by individuals.

Private property involves not only the right of the owner to enjoy all the benefits of his wealth but also the right to say who shall own it after his death; that is, the right of bequest or inheritance. This attribute of private property greatly increases its significance as one of the conditions which control production in the modern world.

Free enterprise. Individual initiative and individual control of enterprise came into being with the relaxation or abandonment of governmental control of the quantities and qualities of goods, of the places where goods could be sold, of prices and other terms of exchange, and of production and trade generally. Freed from the restrictions of government control, men of judgment, initiative, and daring took advantage of the opportunity to plan and operate business enterprises, induced by the prospect of personal gain.

Thus in the United States of America today, it is generally not the Congress, the state legislature, or the city or town authorities that say what sorts of goods shall be produced and what quantities of each, where they shall be marketed, and at what prices sold. Such matters are generally left to the personal judgment and free initiative of the men who direct modern industry: the farmers and miners, manufacturers and merchants, bankers and brokers.

Reaction from laissez faire: General. But we must not conclude from this broad characterization that the principles of *laissez faire* are today in complete possession of the field. The greatest development of economic freedom doubtless came in England and America and in certain of the British colonies and dominions. But nowhere

was *laissez faire* ever completely in operation. Always there have been some governmental controls that would not be countenanced by extreme *laissez faire*. Moreover, there has been going on for a considerable time a significant reaction against the extremes of *laissez faire*. Examination of recent economic history discloses that this reaction has resulted from two fairly distinct motives.

Motives: (a) To make freedom work. In the first place, as time went on it became evident that complete economic freedom would not work. Consider, for example, the history of factory development in England.

The new spirit of freedom, under the doctrine of *laissez faire,* was instrumental in bringing about the Industrial Revolution, which got under way toward the end of the eighteenth century. Industry began to shift from the workers' homes to the factories. Conditions in many of the early factories were far from satisfactory according to modern English and American standards of health, safety, and morals. It is not to be inferred that this was a deterioration in the condition of the common people. Conditions in the factories were on the whole probably superior to what had prevailed in the homes from which the workers came.

Gradually, public attention became focused upon factory conditions, both because of a growing attitude of social responsibility, attending the advancement in national wealth and well-being, and also because the concentration in the factories invited attention which was not attracted by similar or worse conditions in the separate homes.

In the first enthusiasm for *laissez faire,* it had been assumed that under the regime of freedom, such problems would solve themselves. The theory was that each employer was free to hire whom he pleased, and each laborer was free to work or not to work for any particular employer. If the sanitary conditions of a factory were too bad, no one had to work there unless he chose. This employer could not hire workers, since they would prefer to work in another factory where conditions were better. Thus the competition of employers would ensure satisfactory working conditions.

Actually it did not work out this way. The employees, because of inability or inertia, did not gain for themselves factory conditions acceptable to an awakening public opinion. In consequence there started a long history of factory legislation, progressively restricting the freedom of the employer to determine the working conditions

in his factory. Laws required proper lighting, especially of stairways, proper ventilation, provision of adequate toilet facilities, handrails on stairs, guards on machinery to protect workers from accident. Hours of labor were restricted for women and children, and employment of children below a certain age was prohibited.

We have taken the case of factory legislation simply as an example of the many ways in which the extreme freedom of *laissez faire* has had to be curtailed. The important point here is that this movement, though a reaction against *laissez faire,* was not a reaction against freedom. On the contrary, we have learned that freedom itself needs legal protection; otherwise A, in the exercise of his complete freedom, may interfere with the freedom of B.

In the former "horse-and-buggy days," everyone was almost completely free to drive as he chose upon the public highways. In the modern days of the automobile, we have no wish to prevent the utmost possible freedom in the use of the highways. But we have had to recognize that, paradoxically, if all automobile users were allowed complete freedom, no one would be free in the use of the highways. We have had to restrict the highways to licensed operators; we have set speed limits; we have restricted the size and weight of trucks; we have established traffic lights at road intersections; and we have crystallized by law and police enforcement the rules of the road, which in the horse-and-buggy days were generally left to their own enforcement. Were it not for all these restrictions upon the freedom of other motorists, the "freedom" of each one of us to use the public highway would be largely an illusion.

Thus this motive back of the reaction from *laissez faire* is not to turn our backs upon economic freedom, but rather to make economic freedom work. The concession to government authority is the price we have to pay for real and practical freedom. So long as the system of economic freedom is desired by man, and human nature is as it is, there will always have to be such restrictions upon extreme freedom, and progress and the passage of time will lead continually to new forms of control, along with alteration or repeal of former restraints. The present concept of economic freedom includes, and is not contradicted by, such controls.

(b) To displace free enterprise. The other motive behind the reaction from *laissez faire* represents an attack upon the basic principle of free enterprise itself. It springs from distrust of the economic and social results of economic freedom and from a belief

in the superiority of the conscious or authoritarian control of production.

This is the philosophy of the socialists and the communists. While these two groups differ on certain points, and their attitude toward each other is generally characterized by distrust and bitter hostility, they are in agreement in their opposition to economic freedom. In fact they are not content even with full governmental regulation of production in the hands of private persons.

What they demand is the abolition of private ownership of the instruments of production. All the farms, mines, and other lands would be owned by the government, as well as all the factories and their machinery, the railroads and their rolling stock, automobiles (at least those used in production), ships, and in short all land and all capital. Production would not be merely controlled but would be directly conducted by government through the ownership and operation of all the instruments of production and the control of all labor as employees of the state.

Economic freedom has been attacked also by fascism, a doctrine which arose in the interval between the First and Second World Wars and quickly gained full command in Italy and Germany. It is sometimes erroneously supposed that communism and fascism represent the two opposite poles of economic doctrine or policy. Actually in most fundamental respects there is little difference between them. The principal difference is that fascism is not so much interested in government ownership of land and capital. It accomplishes its purposes by the extreme of government control and regulation of privately-owned enterprise. Prices are fixed, wages controlled, business concerns are told what to produce, and how much, and where and to whom to market the product. Government officials are in controlling positions on the boards of directors and among the officers of business corporations and labor unions. While private ownership may continue in name, it is only a shell, not the substance, and true economic freedom is nonexistent.

Another governmental system which represents rejection of, if not reaction from, the principles of economic freedom, is the military dictatorship. Japan, until her defeat in the Second World War, furnished an example. The combination of long-standing feudal institutions with the military dictatorship of more recent times in Japan has always been a bar to economic freedom and free enterprise. Spain represents another example. Here an economic

system largely feudal in character came, in the spring of 1936, under the control of a socialistic regime, which was in turn overthrown in the fall of 1939 by General Franco, who thereafter governed the country by virtue of military power. Military dictatorships are more or less prevalent among the nations of Central and South America.

The present area of freedom. In the years preceding and immediately after the Second World War, socialist and communist sentiment was spreading in many parts of the world. The disintegrating effects of war aided the propagation of such sentiment. It provided also fertile ground for the rise and growth of military dictatorships.

At the present time the institutions of economic freedom are in operation in relatively few countries. We must of course exclude from the area of freedom the uncivilized nations which occupy much of Africa and the Pacific islands, because they are not capable of supporting a system of free enterprise. Other large areas of ancient civilization but arrested development—India and China are examples—do not provide an environment hospitable to economic freedom. What is more significant for our purposes is that, even among nations of advanced civilization and highly developed economic activity, free enterprise is rather the exception than the prevailing rule. Moreover, there has been a tendency in recent years for some such countries to abandon the system of free enterprise previously established in favor of some form of authoritarian control.

In view of the rapid changes now taking place in the political and economic structures of many countries, it is impracticable to draw up a complete and definite list of countries which should be included in the area of economic freedom. We know that such a list would include the United States, Canada, and the Union of South Africa, and perhaps Australia. In Europe only Switzerland and one or two other small countries could be included with certainty; in Latin America not more than one or two countries.

It is uncertain how Great Britain should be classified. The victory of the Labor Party in 1945 was definitely a verdict for socialism in an election which presented a clear-cut issue of free enterprise *vs.* "planned economy." On the face of the matter, therefore, it would appear that Great Britain has withdrawn from the area of economic freedom. But the decision has been taken so recently, and the British people are so traditionally averse to abrupt changes of

national policy, that it is by no means certain how things will turn out.

Viewpoint of this book. In spite of the fact that political and economic freedom prevails today in only the minor part of the world—whether measured in terms of area or of population—the primary purpose of this book is to study the economic facts and principles of the free economy.

This does not imply unawareness of the importance of other economic systems, based upon totalitarian control of economic activity, or of the interest that attaches to their study. Nor do we have any intention to beg the question as to the relative superiority of free enterprise and government control. In the last chapter of this book we shall devote some considerable space to the study of some of these other systems, particularly socialism, communism, and fascism. At that point we shall seek to make some comparison of the economic consequences of different systems and their respective effects upon human welfare.

This book, however, is frankly addressed to American students. Its basic purpose is to aid them in acquiring knowledge of the economic environment in which they live. And the economic system of the United States of America always has been and still is, by and large, based upon economic freedom and free enterprise. It is the functioning of this system and its economic consequences which impinge most directly upon the daily lives of the American people. Hence the viewpoint of this book.

The free enterprise system: The motive of productive effort. The goal of all industry is production, in order that human wants may be satisfied. In the modern economy, each producer is, under the institution of personal and economic freedom, at liberty to produce whatever he chooses, and the institution of private property permits him to have and dispose of what he produces. Thus is the desire for the satisfaction of wants liberated, to become the motive force behind production as it could never be without freedom and private property. The slave may cherish wants without result in increased application and diligence. He does as little as he can without incurring the wrath of the overseer, because he knows that there is small connection between his efforts and the satisfaction of his wants. A free man, on the contrary, has the right to enjoy the fruits of his toil, and he is spurred on to efforts more or less ener-

getic and effective, according to the urgency of his wants and his productive capacity.

But as we have seen, modern society is organized upon a coöperative basis, which means, in the first place, that one does not produce the things that will satisfy his own wants, but must first exchange his own product for goods produced by others, and second, that production must submit to an extremely complex organization controlled by certain individuals, the entrepreneurs.

For the great mass of the workers, the connection between effort and satisfaction of wants, though it still exists, becomes less direct and less obvious. Nevertheless the basic motive remains; the wage earner is aware that the amount of his wages depends, in some measure at least, upon the efficiency of his productive efforts, and the entrepreneur thinks of the profits of his business.

The modern motive to productive effort is thus frankly self-regarding, and modern industry proceeds along the lines of competition, which means in general that each is acting independently in his own interest without primary regard to the interests of others.

Control of production by price. Such being the motive to production, and its direction being in the hands of the entrepreneurs, what is it that guides the judgment of the latter and tells them what goods to produce, what quantities of each, where to send them, and to whom to sell them?

The answer is price. Prices determine profits or losses and so invite or repel the entrepreneur as he casts his eye over all the possibilities in the field of production. Wages, the price of labor, turn laborers from one industry to another. Price directs the flow of savings into new capital equipment for this industry or that. Price determines the amount of iron ore that will be mined, the amount of wheat that will be raised, whether wheat will be sent to one country or another. Price is so omnipresent and so subtle a force that few are fully aware of its significance or realize how completely it controls our economic life.

A simple illustration will make this clear. Ask any farmer what he would do if he knew or firmly believed that the price of wheat were going to double without a proportionate rise in agricultural costs. Having every reason to anticipate greater profits, he would undoubtedly try to grow more wheat per acre and plant some acreage to wheat that had previously been devoted to another crop or lain fallow. In his efforts to produce more wheat he might bid for

laborers, offering them higher wages to induce them to leave their present jobs, and he might purchase new machinery, so stimulating the manufacture of farm machinery and directing capital into that line of industry. Conversely a fall in the price of wheat, threatening lower profits or even losses, would reduce its production, might lower farm wages, and would tend to check the flow of capital into farm machinery.

The American farmers rejoiced in the high prices of the First World War. Profits were large and production expanded enormously to make up for the interruption to the flow of wheat from Russia, Australia, and the combatant countries. The return of peace brought with it an approach to normal conditions. Australian wheat could get to the market, and France could devote capital and labor to raising wheat. The price of wheat dropped, and American farmers found themselves in a desperate plight. The explanation was simple; the world production of wheat was too great to be sold at a profit by all of the farmers. The automatic check was working in the direction of decreased production. This was obviously painful to the farmers after their period of prosperity, but it was the only way to raise the price to a remunerative level.

Guiding production to meet consumers' wants. Price is the medium through which the consumer makes his voice heard, telling those whose judgment directs industry how much of each good to produce and whether any particular good should be produced at all. If someone is willing to pay a sufficient price, anything conceivable will be produced, from bread and coats to diamond shoe buckles. If for something else no one is willing to pay a price from which the producer can foresee a profit, then that thing will not be made. There is no production of the old-fashioned spinning wheels, or of clothing of a bygone style, or of gold horseshoes, because no one will pay the necessary price.

When government authority was displaced by free enterprise under the control of price as the directing force back of production, the control of production was actually given over to the consumers, or more specifically to those who have the wherewithal to pay for the things they want. Production is thus directed, not toward the maximum of things which someone may consider most beneficial to mankind on the basis of some ethical or moral standard, but toward the things which are wanted by those who can pay for them.

Allocating the factors of production among competing uses. Production of anything requires the services of the three factors of production: land, labor, and capital. These factors are strictly limited in quantity; yet the products mankind could use to satisfy his wants are unlimited. With all these products crying to be produced, so to speak, it is price that determines where the limited factors of production shall be employed. High prices for the products in any particular line attract the attention of the entrepreneurs, and they direct land, labor, and capital into that branch of production. Falling prices, with diminishing chance of profits, are the automatic check to overproduction. Prices of raw materials and wages govern costs and have the opposite effects, tending to stimulate production when low, and putting on the brakes when high.

Rationing supplies of scarce goods. Price operates also to ration the supplies of scarce goods. Where an unfavorable growing season results in a short crop of peaches, for example, what determines who will enjoy the scarce fruit? The answer is price. In a free market the price of peaches will begin to rise. Some of the consumers, who would have bought peaches at the previous price, now lose interest. The group increases as the price goes on ascending. Finally an equilibrium is reached at which the stock of peaches just goes around among those who are willing to pay the price.

Here is a smooth-working automatic rationing system, which sees that the scarce good is distributed among those whose demand for it, as measured by their willingness to pay the higher price, is the greatest. This automatic rationing system is in sharp contrast with the governmental rationing with which all Americans became familiar during the Second World War. There was then not a free market; prices were set by the government.

For the most part the purpose of official rationing was to hold the prices down below the levels they would have reached when war conditions brought scarcity of certain goods. The automatic rationing system was thus prevented from working. An authoritarian rationing system had to be imposed, with its cumbersome machinery of "priorities," "rationing boards," "sugar cards," "butter points," "gasoline rations," etc.

When official price-fixing prevails, without an official rationing system, as was the condition for a time after the Second World War, the situation is even more chaotic. Then in the absence of any rationing, automatic or official, we had the queues of housewives

standing hopefully in line at the grocery store in quest of butter, or bread, or meat; the ladies following every rumor regarding the possibility of getting a pair of nylons; storekeepers subjected to various pressures for special favors; and finally the "black market," where price struggles, in spite of legal prohibition, to perform its normal rationing function.

Competitive bargaining. Competitive price is actually determined by bargaining between sellers and buyers, producers and consumers. Each consumer of wheat is striving to purchase his wheat at the lowest possible price, while each seller strives to sell for the highest possible price. The only restraint imposed is that of self-interest. A seller hesitates to be too grasping, lest he lose his trade to another seller; a buyer fears to hold off for a low price too long, lest he lose the opportunity to buy at any price. If the existing stocks appear deficient, this bargaining of sellers and buyers tends to a rise in price and economy in present use, with more held back for future needs. The opposite results follow the fall in price which is induced by abundant present stocks. And finally the high or low price encourages or discourages future production to replace the present stocks.

To sum up, we may say that price (1) guides production into the fields most desired by the consumers, (2) allocates the factors of production among competing uses, and (3) rations scarce supplies of certain goods.

CHAPTER VI

Price and the Market

The problem of price: Price in the free economy. In the preceding chapter the importance of price as the control mechanism of the free economy has been stressed. So important, indeed, is price for this purpose that it would be no exaggeration to say that price is at the very heart of the science of economics.

When division of labor becomes virtually universal, economic life centers about the exchange of goods and services through the medium of money. With comparatively few exceptions, whatever anyone produces must be sold for money, and whatever anyone acquires must come by purchase with money.

A pecuniary system. This pecuniary character of modern life, while in a sense self-evident, is not always given its due place in our thinking. In order to arrive at fundamentals, we are frequently reminded that money is not the aim of economic exertion but is only a medium, serving the purpose of passing goods from producer to consumer with a facility not possible under a régime of barter. The ultimate social purpose of production is not the obtaining of money, but the production of goods to satisfy human wants. And the acquisition of money (except to the traditional miser) is not for its own sake, but simply as a means to the possession of the things that satisfy wants.

These are indeed profound truths. But they are not the terms in which the businessman thinks. To the shoe manufacturer, his factory is operated, not to the end that the community may be comfortably shod, but that profits may accrue to himself. His payment of wages and other costs are not thought of as the means whereby his laborers and others may obtain satisfaction of their wants; they are considered as payments of money, necessary to the conduct of the business. The disposal of his product is thought of in terms of the money brought in, which he regards as the true purpose of his enterprise.

In this sense the shoe manufacturer's goal is not making shoes,

but "making money." This is the businessman's point of view. And the businessman is the centre of the modern economic system.

Just as the businessman thinks in terms of money so do his employees, who receive wages, and so do all of us in our dual capacity of receivers of money incomes and buyers of goods. The modern system is in truth a pecuniary system—a money economy—centred about the exchange of goods and services for money prices.

Prices fixed by statute or custom. It therefore becomes one of the important purposes of economics to learn how prices are determined. There was a time when prices were determined largely by custom and law. People in the Middle Ages talked of "just prices," "fair prices," etc. When a seller sought to exact a price higher than the customary price he at once laid himself open to suspicion. So also of the buyer who sought to compel a price lower than the customary one. In general a price was just if it was as it always had been; the just price was the customary price. Custom was powerful in those days, and the mere force of custom alone had great influence in preventing serious fluctuations in prices. To the aid of custom came law, civil and religious, which undertook, more or less completely, to prescribe what were the fair and just customary prices and to forbid departure therefrom.

But neither law nor custom was able permanently to prevent the price changes which followed in the wake of changed economic conditions. There was some attempt to adapt the law to these changes, and the legal prices were altered from time to time, but the law was unable either to resist the changes or to adapt itself to them. Evidently there were forces at work more powerful than custom and statute law.

The legal determination of price gradually declined and finally ceased to exert an important influence, though the policy has never completely disappeared. In the United States today, legislative authority regulates the rates charged by the railroads for transporting both passengers and freight. Trolley fares, telephone rates, telegraph rates, etc., are subject to legal control. Rates charged by gas companies, electric companies, water companies, and other local public utilities are generally limited by official authority. But the mention of these cases is sufficient to remind us that they are the exceptions and not the rule and that prices in general are evidently not determined by legislation except in times of war.

Custom likewise still plays a considerable part in the control

of price. The fees charged by physicians, lawyers, and other professional men, barbers, bootblacks, etc., are quite generally restricted by custom. But it is evident that, in spite of custom, the prices of all these services do change and that there is at no time universal conformity to the customary schedule. What custom does in these cases is not to determine prices, but rather to act as a steadying force, keeping most prices in line with the customary standard and preventing many great and sudden changes. And after all these prices again are the exceptions, not the rule.

Are sellers or buyers free to fix prices? People sometimes talk as though they thought the retail sellers of goods had complete power over prices. Does not the merchant, we are asked, fix his prices as he sees fit? What does the customer on the other side of the counter have to say about it? He can take the goods or leave them; that is all. But obviously things are not so simple as this for the retail merchant. In the first place, he was a buyer before he became a seller, and the prices he charges are obviously related to the prices he had to pay for the goods. Finally the buyer does have an influence on prices, exercised through his power to refuse to buy.

Is price a mystery? How is it then that prices come to be what they are? We have seen the extraordinary variety of prices. Some at least appear to be entirely freakish. A diamond may sell for a thousand dollars, a loaf of bread for fifteen cents. How is it that the mere luxury is worth over six thousand times as much as a prime necessity of life? A dress suit, worn only on rare occasions, may cost a hundred dollars; a suit for everyday wear, only forty-five. Why does the price of the less necessary thing exceed that of the more necessary? We have seen that prices are constantly changing, some rising while others are falling, and the average of all prices moving steadily downward for years and then as steadily rising.

Is all this complexity a mystery incapable of explanation? Is price merely a matter of chance? Or is it the natural result of causes which may be discovered? Certainly price is of enough importance in human affairs to make the search for its explanation worthwhile if there is promise that the inquiry will be rewarded by success. The investigation of the forces that determine price is the task next before us.

The market: Definition. The word "market," in the vocabulary

of economics, has a somewhat broader content than is given it in ordinary usage. As technically defined, *a market for any good is a place where buyers and sellers exchange that good.*

This definition includes every place where buyers and sellers exchange goods, even though there may be no formal market organization. Thus the New York Stock Exchange is a formal, highly organized market. Less formally organized, though definitely located in one building or area, are the "public markets" maintained in certain cities. Retail markets generally are not organized or definitely located. The retail strawberry market of a city need not be any formally organized market in a special building or place. It may be simply the various retail stores and offices where dealers and consumers sell and buy strawberries.

It is not even necessary that all the buyers and sellers be physically present at the market. They may be in communication with the market by telephone, telegraph, or mail and represented there by agents or brokers. The market is simply the place at which buyers and sellers make their influence felt, at which exchanges are made, the meeting point of the forces that determine price.

Free market. *A free market is a market in which there is no rationing of goods and no official control of prices.* During the Second World War we experienced markets that were not free, where there was "only one to a customer," where most goods could be bought only on presentation of the appropriate "ration points," where the government fixed "price ceilings" on most products, and where purchasers of many articles had to obtain official "priorities." A market with controls such as these is a controlled market. In the following discussion it will be assumed, unless otherwise stated, that the term "market" means a free market as here defined.

Why do people buy and sell? When Mr. Brown purchases a pair of shoes of a particular size, color, and style, he does so, first, because he wants these particular shoes, and, secondly, because in our money economy there is no other practical way to get them. Of course, he might make a pair of shoes for himself. But if he is not a shoemaker but, say a lawyer, for him to do so would be folly. While making himself a pair of shoes he would be losing earnings from his profession worth many times the crude pair of shoes he could make. Moreover, there are experiened shoemakers organized and equipped to provide Mr. Brown with a pair of shoes

of better quality and at much lower cost. The storekeeper is willing to sell Mr. Brown the shoes since, as a merchant, he is in the business of buying and selling shoes for the sake of making a profit for himself.

Common and conflicting interests of buyers and sellers. Understanding of the free market requires recognition of the fact that, in exchanging goods and services, buyers and sellers have both common and conflicting interests in the transactions. Mr. Brown and the storekeeper each seeks an advantage for himself and therefore wants the exchange to take place. But underlying their common interest, each one's interest in the transaction is conditioned by the price. If Mr. Brown is willing to pay the $15 asked for the shoes, he might lose interest if the price were $20, while the lower the price the better Mr. Brown would like it. On the other side, the storekeeper, by quoting a $15 price, is obviously willing to sell at that figure. He probably would not be willing to sell for $10. His interest is to sell as nearly as he can at the price that will yield the maximum profit.

These common and conflicting interests of buyers and sellers appear in a variety of different market situations. It is the purpose of the balance of this chapter to obtain a general and preliminary view of how these interests of buyers and sellers operate to determine price under various market conditions.

Higgling exchange. Farmer Jones intimates to Farmer Smith that he might be interested in buying the latter's bay mare. Farmer Smith indicates his willingness to sell. Here is established an initial common interest in an exchange. But this interest, held by both in common, is strictly conditional and only cautiously admitted by either party.

The conflict between their interests centers in price. Farmer Jones, the buyer, desires to buy the horse at the lowest possible price. Farmer Smith's interest is in obtaining the highest possible price. During the higgling and bargaining over the exchange, each determines for himself just how good a horse the mare is. By successive bids and offers, Jones seeks to discover Smith's lowest selling price and, likewise, Smith seeks to discover Jones's highest buying price. Whether there will be an exchange and, if so, what price will be paid depends, first, upon whether Smith's lowest price is below Jones's highest price, and then upon the relative bargaining shrewdness of the two parties.

Here we have an example of one of the simplest types of market situation: only one article is to be exchanged; there is only one prospective seller and only one prospective buyer; no price is quoted in advance. The result is determined by a process of bargaining, or *higgling,* as the economist likes to call it.

The common auction. Mrs. Pierce decides that she would like to sell her set of antique dining room furniture. She is not certain that there is anyone who will have with her a common interest in the sale. She does not know how the conflict between herself and any prospective buyers as to price will work out. She does want to get the highest price she can get.

So Mrs. Pierce has her furniture put up at auction. The description of the furniture and the time and place of the auction having been publicly announced, those who are interested in the possibility of buying assemble, and the auctioneer calls for a bid. Someone starts by bidding a small amount, say $50. The auctioneer announces this figure and suggests that the furniture is worth much more than that. Someone else, being willing to pay more than that if necessary, bids $75. Then come other bids, each one higher than the last. As the price rises, the less eager bidders, one by one, lose interest and cease bidding. Finally there may be only two bidders left. A has just bid $500. B offers $510. A comes back with $525. The auctioneer looks invitingly at B and listens. When no word comes, the auctioneer, to give all comers one last chance, says: "Going, going . . . gone; to Mr. A for $525." Thus the sale is made.

Such a market process resembles the higgling example in that there is just one article for sale and only one seller. The difference is that there are now several prospective buyers.

The seller has a minimum price below which he would not choose to sell. He usually protects himself against a sale below that price either by having the auctioneer announce it or more usually by having a confederate at the auction instructed to bid the minimum and so obtain the article if no one else should bid as much.

Each of the prospective buyers has his own upper limit, just as in higgling, and each would like to buy the article for a price as much below his maximum as possible. As the bidding proceeds, the price rises successively above the highest prices of the bidders in the reverse order of their eagerness. One by one the bidders

drop out until the article is sold to the most eager bidder, the bidder whose price limit was the highest.

If a seller sets a minimum price, it will be high or low according to his eagerness to sell, and if the minimum price thus exceeds the maximum offer of the most eager buyer no sale will be made.

The seller in an auction does not necessarily receive the highest price that any buyer would pay. He has no assurance that he has received the maximum price that the most eager bidder was willing to pay, but only a price somewhat higher than the maximum of the second most eager bidder.

The auction introduces the phenomenon of competition. There are now several buyers, and they are competing with each other. The result is that the sale is made to the buyer who is most eager as indicated by his offering a higher price than any other competitor.

The buyer's auction: Open bidding. Let us consider another form of exchange. When the government of the Republic of Colombia, which has no printing and engraving establishment of its own, needs some new bank notes, announcement is made of a public bidding at a specified time and place in Bogotá. There accordingly assemble the salesmen of all those printing and engraving firms that may be interested in getting the contract. The official in charge announces the terms of the contract, all except the price, and calls for bids. Someone starts with a good high figure. Someone else bids a lower price. The bidding goes on until finally the contract is awarded to that firm which offers to do the job for the lowest price.

Here we have a form of exchange that is a sort of auction in reverse. Instead of one seller and several buyers, we have here one buyer and several sellers. Allowing for this difference, the economic principles involved are the same as in the ordinary auction. Both the buyer and the seller are initially interested, though not certain that they can reach a mutually acceptable price and so effectuate a sale. The buyer has in mind the highest price that he will pay, which he enforces by reserving the right to reject all bids if none is as low as his top price. Below that he wants the lowest price he can get. Each seller has his minimum price. Each will drop out of the bidding when the bid price goes below his minimum. The contract finally goes to the firm that offers the lowest price, assuming that this is not higher than the buyer's maximum.

Sealed bids. Consummation of the exchange when there is one buyer and several sellers is not always accomplished by the method of open bidding. If a city government desires to have the annual report of its park department printed, the usual procedure is for the city to advertise its requirements and call for "sealed bids" from the printers. Each printing firm that may be interested sends in a bid, stating the price at which it is ready to print the report. When the bids are all in, the official concerned examines them and awards the contract to the company that offered the lowest price.

Here once more we have the initial common interest of buyer and seller, without knowledge as to whether there is going to be an actual transfer since the city reserves the right to reject all bids. As the buyer, the city has in mind the highest price it is willing to pay. Should all bids name prices higher than this, it will reject them all. Below that, it wants the lowest price it can get.

Each seller has his own minimum price, but he would like to get the contract at a price as much as possible above his minimum. Being ignorant of the prices bid by his competitors puts the would-be seller at a certain disadvantage. He will have the best chance of getting the contract if he bids his minimum price. But the higher he can make his bid—and still get the contract--the greater will be his profit. So he has to use his best judgment, seeking to hit upon the highest price, above his minimum, that will still have a good chance of being under the next lowest bidder.

Clearly the method of the sealed bids does not permit quite so close a meeting of minds of buyer and sellers as comes from open bidding. It is nevertheless the method very generally used, by private businessmen and corporations as well as governments, in such matters as selecting a contractor to build a house, or pave a city street, or furnish various kinds of supplies, such as coal, or provide services, such as painting a building.

Price at the buyer's auction. Where an article is purchased by a single buyer as the result of bidding of several prospective sellers, we have competition of sellers. The buyer has a certain maximum price, which will be high or low, according to the degree of his eagerness to buy. So long as any seller bids below this price, there will be a sale, and the height of the price will be related to the eagerness of the sellers. The most eager seller will emerge when he offers a price lower than any other competitor, and it is his eager-

ness to sell that will be the ultimate factor in settling the height of the price.

Just as the seller in the ordinary auction does not necessarily obtain the highest possible price, so the buyer in the buyer's auction is not necessarily assured of the lowest possible price. What he gets is something lower than the minimum of the next lowest bidder; that may still be above the minimum of the final bidder. For the seller to get the lowest possible price would require a final process of higgling with the last bidder. This is not customary in connection with an auction of either sort.

The organized exchange. Still another type of market may be found in the speculative exchanges, such as the Chicago Board of Trade, where there is daily buying and selling of wheat, oats, corn, and other grains, or the Cotton Exchange, or the stock exchanges which have been set up in our larger cities for convenience in dealing in corporation stocks and bonds and the bonds of national, state, and local governments.

We may take the New York Stock Exchange as an example. This is an association—in certain features something like a club—whose members are professional dealers in stocks and bonds, commonly called brokers. The association has an imposing and finely-equipped building in New York City at the corner of Broad and Wall Streets. A large hall is devoted to the actual business of buying and selling, which is carried on each business day from 10:00 A.M. to 3:00 P.M. Scattered about on the "trading floor" of this hall are many "trading posts," something like large circular desks or counters. Each stock or bond which is "listed" for dealings on the Exchange is assigned to one of these posts.

Thus any broker interested in buying or selling, let us say, the stock of the New York Central Railroad Company will go to the particular post to which this company's stock has been assigned. There he can count on finding all the other brokers who at that particular moment are interested in either buying or selling this particular stock.

Bidding may be started by Broker A calling out to the assembled group of brokers that he would like to buy 1,500 shares of N. Y. C. R. R. Co. stock, for which he would like to pay 87⅛, that is $87.125 for each share, the par value being $100. Actually custom has developed a code of words and dumb-show signals by which this announcement is made in less time and fewer words than it takes

to write it. Immediately after A's announcement, Broker B calls out that he will sell 100 shares, but at a price of 87⅞. Broker C then announces that he will buy B's stock for 87¼. He obviously beats the price that A offered. A will then have to decide immediately whether to offer something higher, say 87⅜ or drop out. However, another broker, D, may in the meantime have shouted an offer to sell at 87¾, which beats B's bid.

In this way the various brokers get into the bidding, each announcing the price at which he will either buy or sell. Of course any broker, having bid a certain price, is free to change his price at any time until he has come to agreement with another broker. Just as soon as any would-be seller is satisfied with the price bid by a would-be buyer, and vice versa, these men signal acceptance to each other, and the deal is closed.

The bidding, buying, and selling go on among the brokers present so long as any one is interested or until the Exchange closes. On the occasion of very active trading, the shouting and gesticulating become extremely noisy and excited, and the floor presents a scene which to the uninitiated would look like utter confusion. Yet business goes on rapidly and effectively.

Dealings on the stock exchange present a type of exchange differing from those we have previously examined, in that we now have several parties on both the buying and the selling side. There is competition of both buyers and sellers. The procedure of the exchange offers the perfect method for bringing buyers and sellers together and discovering both whether any two or more have a common interest in exchanging and also the price, if any, at which agreement is possible.

As in other types of exchange, each seller, though he has in mind the lowest price at which he would be willing to sell, is trying to get the highest price he can. He will announce an offer at a price somewhat above his minimum, and by means of the subsequent bidding, in the course of which he may have to lower his own offered price, will eventually either find the broker who will pay him the best price, or learn that no one is ready to give him even his minimum price.

Similarly each buyer has in mind a maximum price. He starts bidding somewhere below that. Eventually he either finds the broker who will sell at the lowest price or discovers that there is no one willing to sell except at a price higher than he will pay.

Price quoted by the seller. The common feature of the four types of exchange which we have so far examined is that buyers and sellers come together without any previous announcing of the price. Price is the principal matter that has to be determined by their higgling or bargaining or bidding.

This fact makes these market situations all different from our earlier example of Mr. Brown and his shoes. When he went to the shoe store, he found the price already announced by the storekeeper—the seller. There was no occasion for higgling over the price or of making or receiving bids. He could either take the shoes at that price or leave them. If the price was above his maximum, he did not buy. If it was below, he made his purchase, with more or less satisfaction on account of the margin below what he would have paid if he had had to.

This form of exchange is obviously typical of the sales in retail stores and by service agencies such as the laundries, cleaning and pressing establishments, gas companies, railroads, etc. It is the sort of exchange by which we acquire the great bulk of all the consumers' goods which serve to satisfy our everyday wants.

In such transactions both the storekeeper and the customer have a common interest in making the exchange. The conflict of interest may not be quite so clear, but it is there. The customer would like to get the article at a low price and he has in mind a maximum price above which he will not go. But he has no opportunity to higgle with the storekeeper in order to get a lower price than is marked. If he thinks the price is too high, he can merely refrain from buying—"take it or leave it."

The storekeeper, on the other hand, has his own minimum price, below which he will not sell. Economists call this the *reservation price.* The storekeeper does not usually quote the price at this minimum, since it is generally to his interest to get a higher price. What price above his minimum he quotes depends upon his judgment as to the willingness of his customers to buy at various prices. But he declines to higgle in order to make a sale to a reluctant customer. If the customer considers the price too high, the storekeeper simply loses that sale.

Price quoted by the buyer. A variation from the last example is to be found in connection with the sale of grain by the American farmers, especially in the central and western states. A farmer, for example, after harvesting his wheat, takes it for sale to the ele-

vator company in his neighborhood. Arrived there, he finds a price quoted at which wheat will be purchased. He has already determined the lowest price at which he will sell. If he finds the elevator's price below this, he has nothing to do but take his wheat home and either hope for a better price later or come down on his own minimum. If the elevator's price is the same as his minimum or higher, he sells, with more or less satisfaction according to the margin above his own minimum.

Here again we have competition of both buyers and sellers. The contrast between this and the ordinary buying at the retail store is that, whereas in the stores the prices are quoted by the seller, leaving the buyer the choice of "taking it or leaving it," in the local wheat markets it is the buyers who quote the price and the sellers who "take it or leave it."

EXERCISES

1. In the so-called Dutch auction it is the custom "to offer property at a price above its value, then gradually lowering the price, till someone accepts it as a purchaser." (*Webster's New International Dictionary,* p. 686.) Compare this method with those described in the text. Show how and to what extent the sellers' and buyers' interests are reconciled in the final price.
2. In the discussion of the commodities traded in the various types of markets described, what do you observe is the individuality or uniformity in the several markets? Would you say there was a market for wheat—or separate markets for Manitoba No. 1 winter wheat, etc.? a market for stock or a market for General Motors Company common stock, etc.?

CHAPTER VII

Price Determination Under Pure Competition

Competition. From the seller's viewpoint, the most significant classification of markets relates to the character and degree of competition which prevails between himself and other sellers. *Competition is the effort of two or more sellers to sell to the same person or persons, each seller acting independently in his own interest without regard to the interests of other sellers.* This is obviously competition of sellers. There may be also competition of buyers, with an analogous definition, but this concept is seldom employed. Where the term competition is used without qualifying words, competition of sellers is to be understood.

In the study of markets and the forces that determine prices, it is essential to distinguish two types of competition, *pure competition* and *imperfect competition.*

Pure competition. Pure competition exists when the following conditions are met:

(1) The product exchanged must be, in the minds of the buyers, identical for all sellers, so that a lower price is the only factor that will lead buyers to prefer one seller to another.

(2) The total number of sellers must be very large.

(3) The amount which each seller can offer for sale must be a very small part of the total quantity offered, so that no change in the amount he offers can have an appreciable effect on the price.

(4) All sellers and buyers must be well informed and keen, and acting with complete independence of each other.

When these conditions are all present in any market, no seller can by himself exercise any control over the price. Since any seller's product is considered by buyers to be just like those of his competitors, and since the buyers are well informed and keen, no one will buy from a seller if he asks a price above that of other sellers. Since each seller offers only a small share of the total of goods for sale, even though he should drop out entirely the buyers

could still get all they wanted from his competitors without any appreciable effect on the price. Moreover no seller has any inducement to ask a lower price than other sellers, since he can sell all his products at the going market price. His own share of the total is so small that any increased quantity of goods he might offer at the market price would have no practical effect on the market.

It is possible, therefore, to combine the several requirements of pure competition in a simple definition as follows: *Pure competition is competition between sellers under such conditions that the individual seller is unable by his own actions appreciably to influence the market price.*

Where pure competition prevails. Pure competition does not exist in three of the market situations previously discussed. Where there is an exchange by bargaining between two individuals—higgling—there is obviously no competition of any sort. In the common auction, though there is competition of buyers, there is no competition of sellers. In the buyers' auction there is competition of sellers, but not pure competition, since the number of sellers is too small, the quantity offered by each is too large a portion of the whole, and each competitor generally offers to sell the entire quantity desired by the buyer.

All the conditions of pure competition are generally met on the organized exchanges, where dealings take place in standard commodities such as wheat and cotton, and in corporation securities and government bonds. Although a producer or dealer may in these markets temporarily influence the price by extraordinarily large offers under very special conditions, generally the price of wheat at the Chicago Board of Trade or the price of General Electric Company stock on the New York Stock Exchange is beyond the control of any individual seller.

Pure competition is to be found also in those markets where the buyers normally quote the price. In such markets where staple agricultural commodities such as wheat and cotton are exchanged, the individual wheat farmers or cotton growers may not appear to be competing with one another. They are selling a uniform standard commodity, and there are so many of them that no one of them is able to influence either the price or the fortunes of his fellows. But these are the very characteristics of pure competition.

Imperfect competition. Imperfect competition, the second of the two major types of competition, exists when any one of the essen-

tial conditions of pure competition is lacking. Thus the products offered by the different individual sellers may be differentiated by one or more peculiarities so that they call forth preferences among the buyers. Although appealing to the same wants of consumers, the product exchanged is not, as in pure competition, identical for all sellers from the point of view of the buyers. Or instead of a very large number of sellers, there may be so few sellers that the amount offered by each comprises, as in the steel industry, a substantial portion of the total quantity of the product offered. Finally there may be market situations where buyers are uninformed and slow in acting to gain their maximum advantage, and sellers, rather than acting independently as in pure competition, are controlled by some form of agreement among themselves.

Automobiles provide a striking illustration of a condition of imperfect competition due to differentiation or nonstandardization of the product. Although Ford, Chevrolet, and Plymouth cars are in competition with each other for the automobile purchasers' dollars, each brand has its individual peculiarities, so that to automobile buyers they are different, even though closely related, products. Rarely do they sell at the same price, but competition between them keeps their respective prices within a relatively narrow range. Some buyers may be, and often are, ready and willing to pay an extra few dollars to obtain the particular brand of automobile they prefer. Under such conditions, sellers do have some, though limited, power to influence price.

These negative conditions may be summed up in the following positive definition: *Imperfect competition is that form of competition in which the individual seller is able by his own actions to exert some influence upon the market price.*

Determination of price on the organized exchange: Price and demand. In addition to its intrinsic interest, what makes the exchange especially pertinent to our present inquiry is that it offers perhaps the simplest and clearest example of the operation of competitive market forces in determining price.

In the organized exchange we note at once an important point of contrast with the higgling and auction markets. In those cases there is just one article or a fixed quantity for sale. All that the person has to decide, with respect to any particular price, is whether to buy or not to buy, to sell or not to sell. At the auction of Mrs. Pierce's furniture, A has to decide only what price he would be

willing to pay or, at any particular moment, whether he would be willing to buy the furniture at a price somewhat above the last bid.

When Mr. Davis goes onto the floor of the stock exchange and, approaching the General Motors post, finds that this stock can be purchased for $70 a share, he is not confronted with the alternative of buying or not buying just one share. He can buy as many as he pleases. What he has to decide is (1) whether at the present price he wants to buy at all and (2) if he wants to buy, how many shares.

The key to the problem of price determination under these circumstances is the well-known principle that eagerness to buy or sell is related to the price. The height of the price at any given time will determine the number of shares that buyers will want to take. If the price is high, some who would have been interested at a lower price will not buy at all, and those who buy will take fewer shares than they would at a lower price.

This well-known relation between price and the quantity that buyers would take can be easily put in the form of a schedule. Let us suppose that we have knowledge that at a certain moment the dealers on the New York Stock Exchange would like to buy 30,000 shares of General Motors stock if they could get it for $72 a share; at a price of $71.50 they would take 40,000 shares; 50,000 shares at $71; and so on. Going the other way, they would take only 20,000 shares if the price were $72.50; 13,000 shares at $73; 10,000 shares at $73.50; or 9,000 shares at $74. This state of mind of the brokers is conveniently shown in the following schedule (extended a bit for a few lower prices).

Prices	Quantities that dealers are ready to buy
$74.00 per share	9,000 shares
73.50 " "	10,000 "
73.00 " "	13,000 "
72.50 " "	20,000 "
72.00 " "	30,000 "
71.50 " "	40,000 "
71.00 " "	50,000 "
70.50 " "	60,000 "
70.00 " "	70,000 "
69.50 " "	80,000 "
69.00 " "	95,000 "
68.50 " "	120,000 "

Economists call such a schedule a *demand schedule,* or simply *demand.* This particular schedule is our assumed demand on the New York Stock Exchange for shares of stock of the General Motors Corporation at, let us say, 11:30 A.M. September 2, 1947.

Graphical representation, a digression. Study of the facts presented in a schedule is generally made easier if the schedule is translated into a curve or graph.

In almost all English and American economic charts, quantities of goods are measured along the horizontal axis, and other magnitudes (prices, costs of production, etc.) along the vertical axis. This is general custom, purely arbitrary, but convenient since it avoids confusion. To make the chart clear, however, it is advisable to print the word "Quantities" along the horizontal axis, and another appropriate word or phrase along the vertical axis.

To construct a graph of our demand schedule, we locate on the chart points representing the respective relation between the possible prices, from $74 to $68.50, and the corresponding numbers of shares that buyers would take. The first line of the schedule indicates that at a price of $74 buyers would take 9,000 shares. To represent this supposition graphically, locate 74 on the vertical scale and proceed in a straight line horizontally to the right until a point is reached exactly above 9,000 on the horizontal scale; or locate first 9,000 on the horizontal scale and then proceed in a straight line vertically upward until a point is reached exactly on a level with 74 on the vertical scale. Either way we arrive at the point marked *a* in Figure 2. This point represents the supposed fact that if the price were $74 buyers would be ready to take 9,000 shares of General Motors stock.

In similar fashion eleven other points (*b, c, d, . . . l*) may be located, representing the eleven other suppositions regarding the numbers of shares of stock that would be taken at various prices as shown in the demand schedule. Twelve points in all are located. They are then connected by a series of straight lines. Thus a "curve," *al,* is obtained, which is known as a *demand curve.*

The curve thus constructed is a demand schedule converted into graphical form. Every significant point on the curve bears a definite relationship both to the horizontal and to the vertical axis. The distance of such a point from the horizontal axis, as measured by the vertical scale, indicates a certain price. The distance of this point to the right of the vertical axis, as measured by the hori-

zontal scale, indicates the number of shares our supposed dealers would be ready to buy at the price indicated. For example, the length of the line *mg* tells us that at a price of $71, buyers would

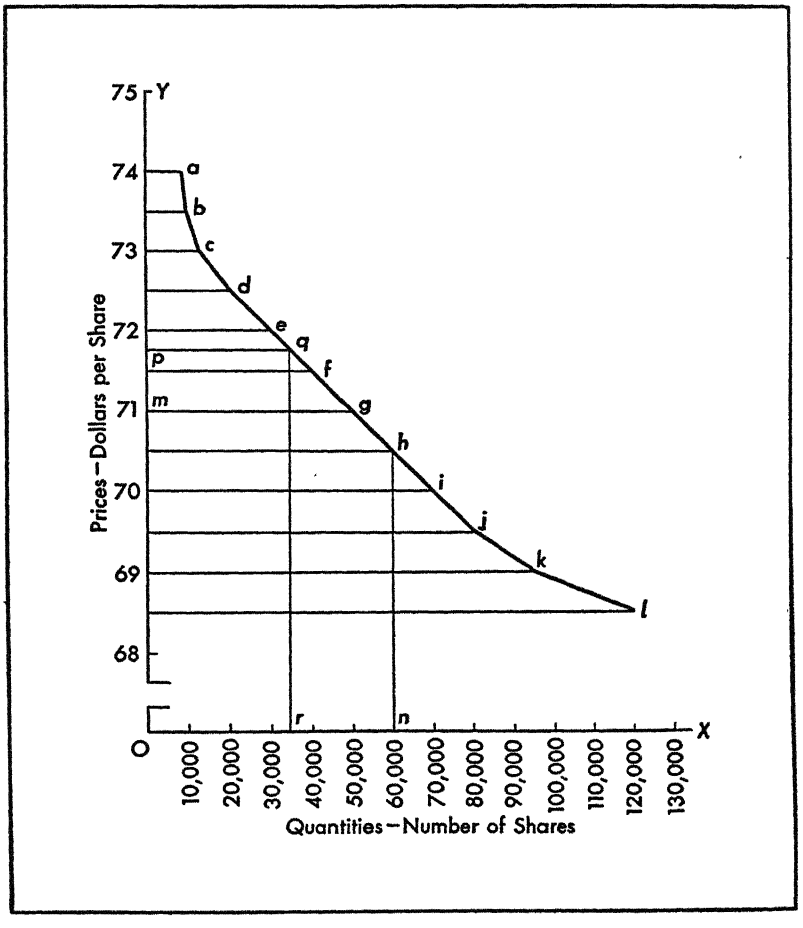


FIG. 2

like to take 50,000 shares. On the other hand, the length of the line *hn* shows that, in order to interest buyers in taking just 60,000 shares, the price would have to be $70.50.

The points *a, b, c, . . . l* represent all the information stated in the schedule. Strictly the diagram should not contain a curve

at all, but merely a series of dots, or points. The points are joined by a curve simply in order to bring out more clearly the relation between price and the number of shares that would be taken by buyers.

However when intermediate numbers between those given on the schedule are possible, we may estimate by interpolation for any given price the quantity which buyers would take, or for any given quantity the price at which buyers would take that quantity. Suppose it is desired to know how many shares would be taken at some price not stated in the schedule, say $71.75. To obtain the number graphically, we have only to locate the point p measuring $71.75 on the vertical axis, draw from p a horizontal line intersecting the curve at q, and drop from q a perpendicular to the horizontal axis intersecting it at r. The distance of r from the origin shows that the number of shares that would be taken at price $71.75 is 35,000.

The simplicity of the illustration used in this elementary discussion of the graphic method does not adequately emphasize the services of the curve as an aid in analysis of statistical data. The student of intricate economic phenomena in the business world finds that so vast and complex is the mass of data with which he must deal that other methods of analysis are inadequate. Of course the curve adds nothing to the evidence of the data from which it is constructed. It does, however, reduce complex data to a simplified form so that inherent truths may be brought to light, thus abbreviating the search for the laws which govern the relations of different factors. Even where other methods are effective, the graphical method is usually serviceable for preliminary and supplementary studies.

Price and supply. While buyers are eager to buy in inverse relation to the height of the price, the eagerness of sellers, on the contrary, is in direct relation to the height of the price. If the price is low, few will want to sell, and most of those who are willing to sell will offer fewer units than if the price were higher. If the price is high, some who would not be interested at the lower price are now ready to sell, and most sellers will offer larger quantities than at a lower price.

This relation between price and the quantity that sellers would offer can be put in the form of a schedule. Returning to our example of the market for the stock of the General Motors Corpora-

tion, we may suppose that the composite frame of mind of the sellers at a certain day and hour is indicated by the following schedule:

Prices	Quantities that dealers are ready to sell
$74.00 per share	118,000 shares
73.50 " "	116,000 "
73.00 " "	110,000 "
72.50 " "	100,000 "
72.00 " "	90,000 "
71.50 " "	80,000 "
71.00 " "	70,000 "
70.50 " "	60,000 "
70.00 " "	50,000 "
69.50 " "	40,000 "
69.00 " "	30,000 "
68.50 " "	6,000 "

Such a schedule is called a *supply schedule,* or simply *supply.* This particular schedule is our assumed supply of stock of the General Motors Corporation at, let us say, 11:30 A.M., September 2, 1947, on the New York Stock Exchange. Following the technique discussed in connection with demand, we construct in Figure 3 (page 108) a "supply curve" representing the foregoing schedule.

Demand and supply together. Assuming that the demand and supply schedules which we have separately considered are in effect at the same time, *i.e.,* at 11:30 A.M. on September 2, 1947, let us plot them both on the same graph, as in Figure 3. The curve *dd* is the demand curve, identical with the curve *al* in Figure 2. The curve *ss* is the supply curve, the graphical representation of the supply schedule.

The particular numerical quantities here set forth are not significant. Their relative magnitudes conform to the well-known relations between price and the quantities that buyers would be willing to take or sellers would offer. The particular quantities are used simply for purposes of illustration.

The curves of demand and supply intersect at the point *p*. The height of this point, *pb,* represents the same price, and its distance from *OY, ap,* the same quantity, on both curves. As a point on the curve *dd,* the point *p* indicates that at a price of $70.50 dealers on the Stock Exchange are able and willing to buy 60,000 shares of General Motors stock. As a point on the curve *ss, p* indicates that at a price of $70.50 other dealers at the same time are able and

willing to sell 60,000 shares. That is, at the price of $70.50 quantities demanded and offered are the same.

This intersection point *p* is of particular significance under our assumed conditions of demand and supply. Suppose that on the

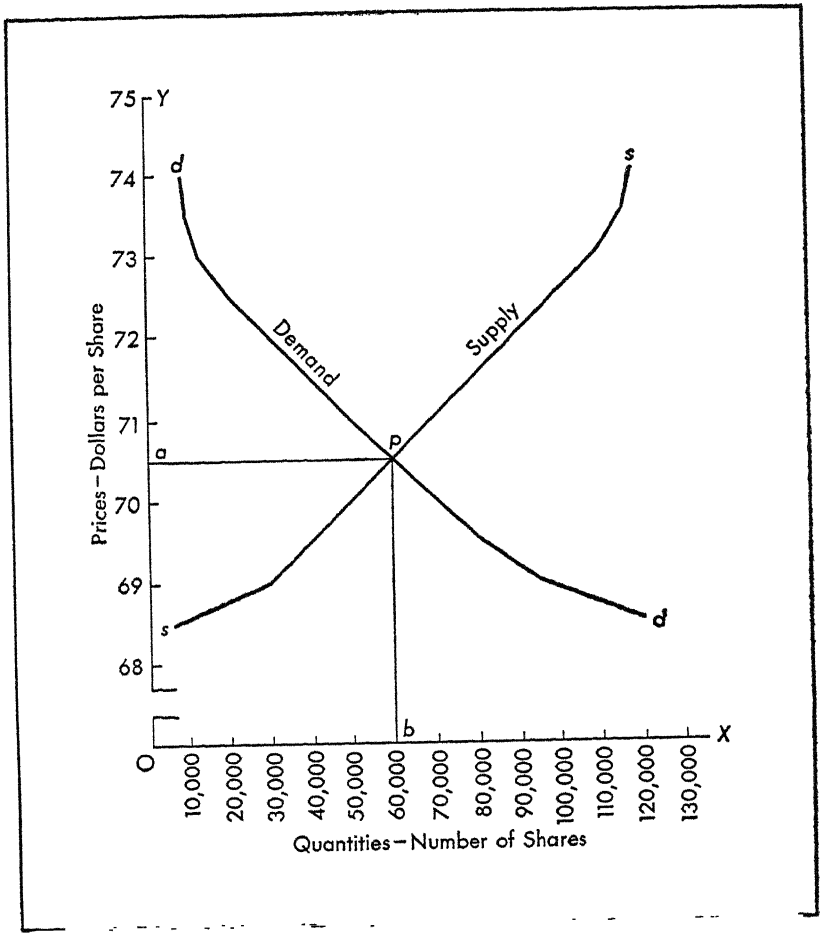


FIG. 3

morning in question business at the General Motors post is started by some sales at a price of $71, as measured by the line *Oe*, in Figure 4. This graph shows that at this price, the buyers altogether are ready to take 50,000 shares, as measured by the line *eh*. Sellers

on the other hand are ready to sell 70,000 shares, as measured by the line *ei.* If this price should hold, the buyers would get all they wanted (at $71), while the sellers would be left with 20,000 shares which they would have liked to sell but which the buyers would

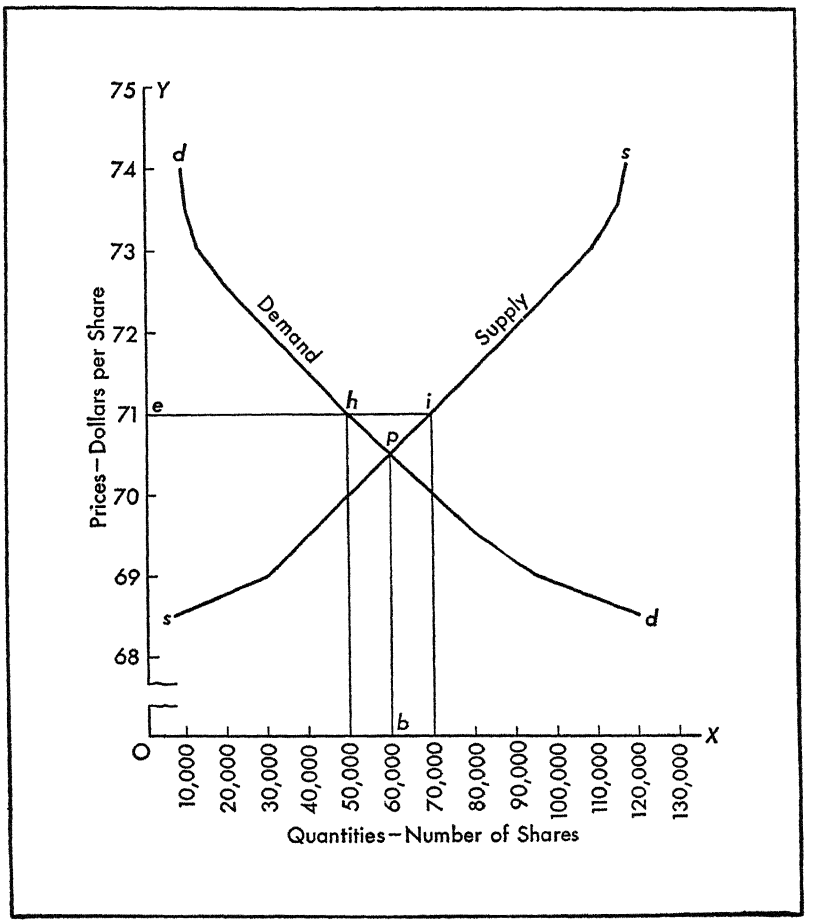


FIG. 4

not take. This is indicated by the line *hi,* the difference between what the sellers would like to sell, *ei,* and what the buyers will take, *eh.*

Under these circumstances, the market is not in equilibrium. Since this is a market of pure competition, in which there are many

independent buyers and sellers, each well informed as to present conditions of demand and supply, such a condition cannot long persist. The dealers will soon recognize the situation and act accordingly.

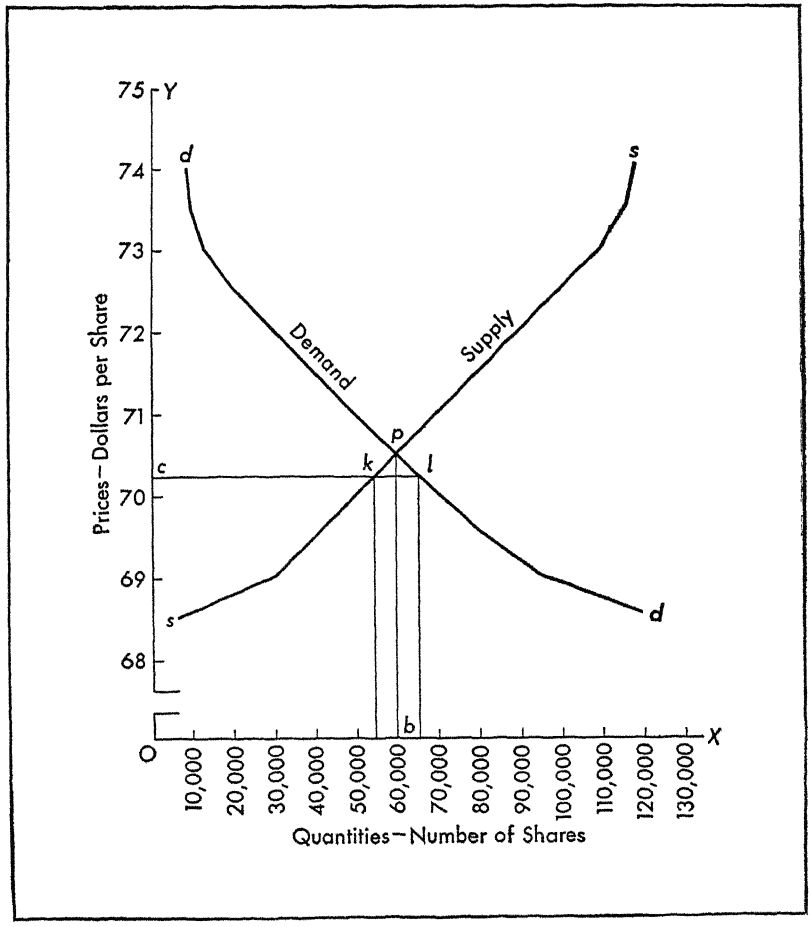


FIG. 5

Sellers will rush in eagerly to sell at the favorable price. Recognizing that they will not all be able to sell all they would like, they will begin to ask slightly lower prices in order to make sure of selling.

Buyers, sensing the situation, will offer lower prices and hold back in the expectation of lower prices.

As this goes on, the price will decline by successive bids and offers, till it finally reaches *bp*.[1] Here there is equilibrium, since there is no longer any discrepancy between the quantity that buyers are ready to take and the quantity that sellers will offer.

In like manner if the market opens with some sales at a price of $70.25, as measured by the line *Oc* in Figure 5, buyers are ready to take 65,000 shares, as measured by the line *cl*, and sellers will part with only 55,000 shares, as measured by the line *ck*. If this price should hold, the sellers could sell all they wanted, but the buyers would fail to get as many as they wanted by 10,000 shares, as indicated by the line *kl*.

Here again is a condition of disequilibrium. Buyers will begin to bid slightly higher prices and sellers will offer to sell at higher prices. The price will thus rise step by step, assuming demand and supply remain unchanged, until it arrives at *bp*, where there is no longer any difference between the quantities offered and demanded.

Equilibrium of price and quantity. Clearing the market. The price *bp* is therefore *the price* of General Motors shares at this particular time. Any other price is not in equilibrium and sets in operation forces that tend to bring about equilibrium at the price *bp*. These forces determine at the same time the quantity exchanged; *i.e.*, 60,000 shares, as measured by the line, *ap* or *Ob* in Figure 3. This is the quantity that corresponds to the price of $70.50, *bp*, on both the demand curve and the supply curve. At that price all sellers are able to dispose of all the shares they want to sell (at that price) and all buyers are able to get as many shares as they wish to buy (at that price). No dealer who is willing to take $70.50 will be left with shares he would like to sell, and no dealer who is willing to pay $70.50 will fail to get as many shares as he wants. This price and no other, under the assumptions, *clears the market.*

The laws of price equilibrium. The conclusions at which we have now arrived may be stated in the more formal language of economic laws, as follows:

I. *If the quantity of any good which people are ready to buy at*

[1] For this conclusion we assume that, during the process, the conditions of demand and supply have not changed. If they have, the equilibrium would be at a different price and quantity.

the prevailing price is greater than the quantity offered at that price, the price will rise.

II. *If the quantity of any good which people are ready to buy at the prevailing price is less than the quantity offered at that price, the price will fall.*

III. *At equilibrium, the price of any good and the quantity of the good exchanged are determined by the two forces of demand and supply, at that point at which the quantity which people are ready to buy is equal to the quantity which people are ready to offer for sale.*

Alternative concepts of demand and supply: Two possible meanings of demand. The whole investigation of the laws of price is in danger of being confused by the fact that the term "demand" is used with two different meanings in popular speech and writing, and even in much scientific writing. For example we hear "there is a heavy demand for wheat"; "the demand for high shoes is decreasing rapidly"; "the demand for automobiles is unusually strong this spring." In all such expressions it is evident that the word "demand" refers to the whole *demand schedule* (or curve).

The meaning is that at any given price the quantity that buyers would choose to take is large or small, greater or less, etc. In Figure 6, the curve $d'd'$ represents a heavy demand as compared with the curve dd, since whatever price is chosen, the quantity that would be taken according to $d'd'$ is greater than according to dd. Thus if dd represents the demand for coal at a given time, a price of $15 would on this day induce buyers to take 100 tons; if on some other day the demand were as shown by the curve $d'd'$, this means that buyers would on that day take 250 tons at $15 a ton. The same notion of demand is likewise in mind when such expressions as these are used: "The price of anything depends on supply and demand." "The greater the demand, the higher the price."

On the other hand, one meets such statements as these: "An increase in price causes a decrease in demand." "The higher the price, the smaller the demand." These statements are in direct conflict with the last quotation in the previous paragraph. Something must be wrong when people say, "The greater the demand, the higher the price," and in the next breath, "The higher the price, the smaller the demand."

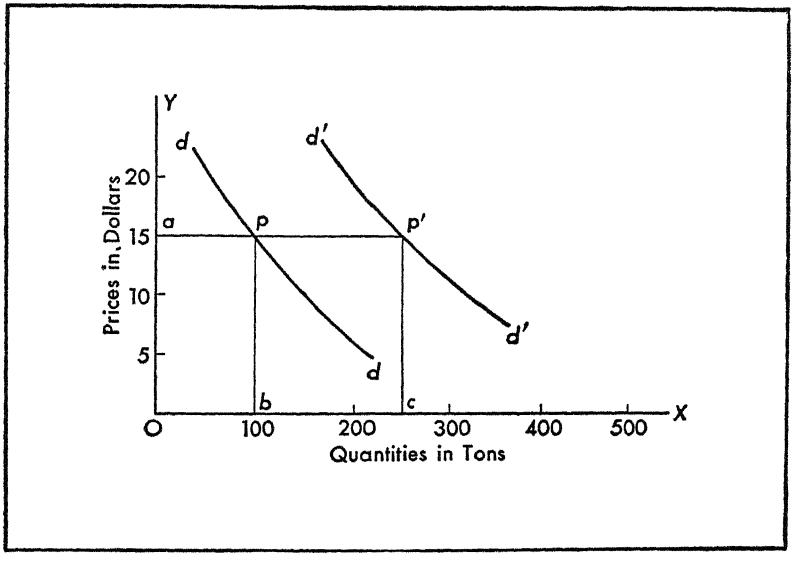


FIG. 6

The trouble comes from shifting the meaning of the word "demand." In the second expression demand does not mean the whole demand schedule or curve. It means rather *the quantity people are ready to buy at some particular price.* In Figure 6, according to the curve *dd,* at a price of $15 buyers would take 100 tons of coal. This quantity, 100 tons, is *the demand for coal at $15,* when the word is used in this second sense. It is represented by the line *ap* or the line *Ob.*

Suppose on the day the curve *dd* is in effect, someone says the demand for coal would be less if the price were raised above $15, or the demand would be greater if the price were less. He does not mean a change in the whole demand schedule or curve, but rather a shift from one to another of a series of simultaneous possibilities. The ability and willingness to buy any other quantity at any other price, as shown by the curve, exists at the same time as the ability and willingness to buy 100 tons at $15.

In graphical terms this talk of a decrease or increase in demand is described by a movement along the curve, locating a different point, whose distance from the *y* axis is greater or smaller than before. When used in this sense we can never think of demand

except in reference to a *particular price,* stated or implied. We cannot speak simply of the demand for cotton. We must say: the demand for cotton at fifteen cents a pound, or at some other price.

Choice of a definition of demand. It is idle to argue about definitions, since there is no absolute test of the correctness of a definition. Either of the two possible meanings of the term "demand" might be chosen. Both are equally favored in the loose usage of popular speech and writing. We violate popular usage to some extent when we choose to make exclusive use of either meaning, but that is inevitable. We should select, therefore, the meaning which will best serve the purpose of the scientific investigation of economic laws.

The central problem in economics is concerned with the laws which determine value and price. In the study of this problem, while it would be possible to make fairly satisfactory progress with either definition of demand, the advantage is decidedly in favor of the first meaning. We have therefore accepted the definition: *The demand for any good is a schedule of the respective quantities of that good which people are ready to buy at all possible prices.*

Two possible meanings of supply. Choice of a definition. The term "supply" is commonly used in two senses, corresponding exactly to the two meanings which are given to demand. When people say, "The supply of cotton is short this year," or "The wheat supply is greater than ever before," they are using the term as the *whole schedule* of quantities that people are ready to offer at various prices. They mean, for example, that at whatever price bid the quantity of cotton which sellers are ready to offer is small, relative to the quantities they are ordinarily ready to offer. The term is used also in this sense when it is said: "The price of anything depends on demand and supply." "The greater the supply, the lower the price."

However such expressions as these will frequently occur: "An increase in price causes an increase in supply." "The supply depends on the price." "The lower the price, the smaller the supply." There is here the same confusion as results from the two meanings of demand. For example, the last quotation above is clearly in flat contradiction with the last quotation of the previous paragraph. Again, the trouble is due to shifting the meaning of the term "supply." In the sentences quoted in this paragraph, "supply"

does not mean a schedule of quantities that would be offered at various prices; it means *the quantity people are ready to offer at some particular price.*

All of this corresponds evidently to the two meanings of the term "demand." Here also satisfactory progress might be made with either definition, but the advantage is with that which regards supply as the whole schedule rather than a particular quantity offered. Hence we have adopted this definition: *The supply of any good is a schedule of the respective quantities of that good which people are ready to offer for sale at all possible prices.*

The importance of consistency. While satisfactory results might come from the consistent employment of either pair of possible definitions of demand and supply, confusion inevitably results from a vague use or from a mingling of the two meanings which might be attached to each of these terms. Equally essential is it that the definitions chosen for demand and supply be in harmony with each other. We cannot define demand, for example, as a schedule of quantities that would be bought, while defining supply as the quantity that would be offered at a particular price. We have chosen the schedule sense for both terms, and these definitions will be adhered to consistently throughout this book.

The causes of price change: In general. Our excursion into the theory of price has thus far shown us how prices of goods in a market characterized by pure competition are determined, as of a particular time, by demand and supply. Prices in the practical world of today are however seldom stationary but are generally subject to change, more or less frequent. It is not enough therefore to know how a certain competitive price may be established in a static market. We must also understand the mechanism by which prices rise and fall under the dynamic conditions of modern life.

For the sake of a concrete example, let us assume that we are investigating the price of fresh-laid eggs in the retail market of a small-sized city, which we call T. We assume that on a certain day, October 15, 1946, there were no changes in other prices, but that the price of eggs at the beginning of the day was somewhat uncertain. The demand for eggs and the supply of eggs were as shown in the schedules on page 116. The relations between the quantities and the prices in these schedules are the same as those developed in our previous illustration relating to General Motors

stock. The exact numerical quantities are of course not significant; they are used only for purposes of illustration.

RETAIL EGG MARKET OF T, OCTOBER 15, 1946

DEMAND		SUPPLY	
Prices	Quantities people were ready to buy	Prices	Quantities people were ready to offer
90 cents	140 dozen	90 cents	320 dozen
85 cents	160 dozen	85 cents	310 dozen
80 cents	180 dozen	80 cents	300 dozen
75 cents	200 dozen	75 cents	290 dozen
70 cents	220 dozen	70 cents	280 dozen
65 cents	240 dozen	65 cents	270 dozen
60 cents	260 dozen	60 cents	260 dozen
55 cents	280 dozen	55 cents	250 dozen
50 cents	300 dozen	50 cents	240 dozen
45 cents	320 dozen	45 cents	230 dozen
40 cents	340 dozen	40 cents	220 dozen
35 cents	360 dozen	35 cents	210 dozen
30 cents	380 dozen	30 cents	200 dozen

Now let us plot on one diagram two curves corresponding respectively to these two schedules. The result is shown in Figure 7; *dd* is the demand curve; *ss*, the supply curve.

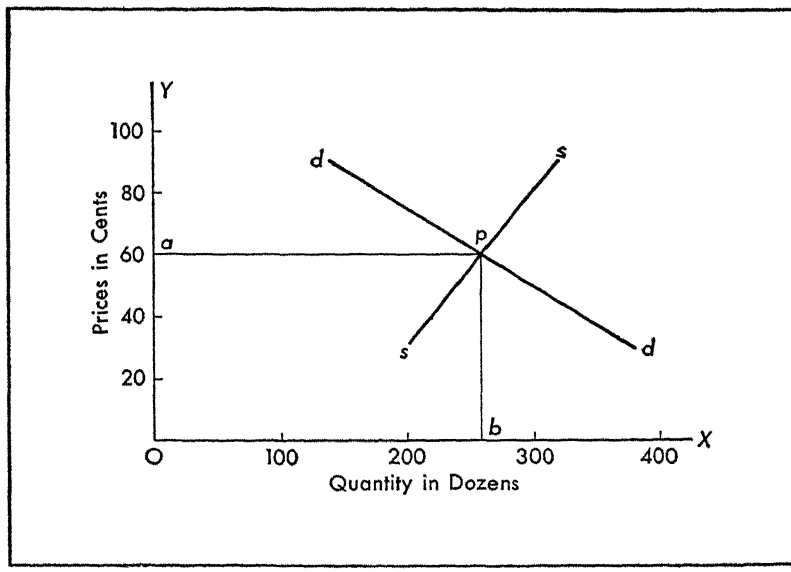


FIG. 7

The curves of demand and supply intersect at the point *p*. The height of this point, *pb*, represents the same price, and its distance from *OY* the same quantity, on both curves. The point *p*, as a point on the curve *dd*, indicates that at a price of 60 cents people in the city of T were able and willing to buy 260 dozen eggs; as a point on the curve *ss*, *p* indicates that at a price of 60 cents other people in the same city were able and willing to sell 260 dozen eggs. At the price of 60 cents quantities demanded and offered were the same.

No other equilibrium price than *pb* is possible with these particular curves of demand and supply, since *p* is the only point at which the curves can intersect. A different point of intersection could be obtained only by a shift in one or both of the curves. With the path thus indicated our present inquiry is quite simple. We shall find the graphical aids especially useful in this part of our investigation.

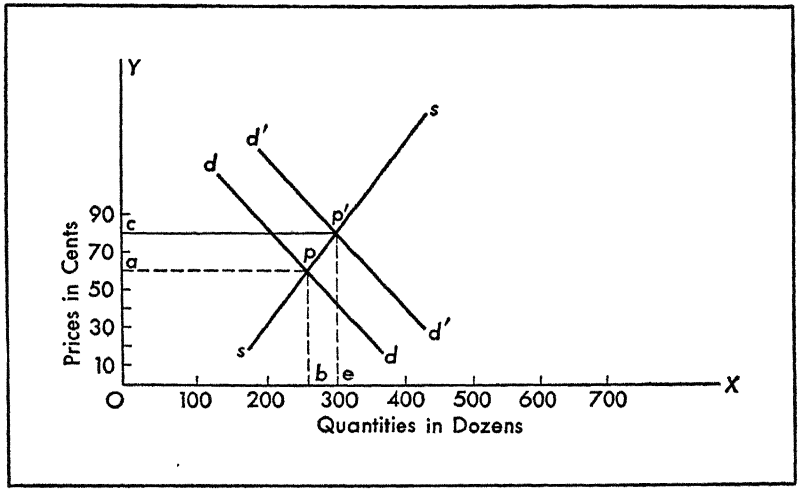


FIG. 8

Let us consider an increase in demand. This means that buyers have become more eager. In terms of the definition of demand, it means that people are able and willing to buy greater quantities at all possible prices. Graphically this situation is represented by a new demand curve, each point on which is farther away from

the y axis than each corresponding point on the old curve; in other words, the demand curve has been shifted to the right.

Let us assume that, as cold weather comes on, the people of T are inclined to consume more eggs; in other words, the demand increases. Let us suppose that the demand on December 15 would be correctly represented by the curve $d'd'$ in Figure 8. Assuming further, for the sake of simplicity, that there has been no change in the conditions of supply, the curve ss will serve to represent the supply both on October 15 and on December 15. (Compare Figure 7 on page 116.) The new demand curve intersects the supply curve at the point p', farther to the right and farther up than the point p, which determined the price and the quantity exchanged on the previous date. The price is now 80 cents, and the quantity of eggs exchanged is 300 dozen, an increase in both magnitudes since the earlier date. We conclude then that, under conditions such as these, an increase in demand tends to cause an increase in price and an increase in the quantity exchanged.

By employing similar reasoning the reader can easily round out this analysis by demonstrating and illustrating these three additional propositions: (1) a decrease in demand tends to cause a decrease in price and in the quantity exchanged; (2) an increase in supply tends to cause a decrease in price and an increase in the quantity exchanged; and (3) a decrease in supply tends to cause an increase in price and a decrease in the quantity exchanged.

The foregoing generalizations have been stated as tendencies, meaning that the results indicated will follow the respective causes unless some other cause interferes—that is, if other things are equal. Thus we have noted the effects of changes of demand or of supply, the other factor remaining unchanged.

But it is quite possible and normal that changes in both factors may occur at the same time. The hens may lay more eggs, thus increasing the supply, at the same time that the approach of hot weather causes a decline in demand. In such cases the effect is the resultant of the two influences, operating jointly in accordance with the generalizations previously stated.

It will be readily apparent that in some cases the two influences may reinforce each other while in other cases they may neutralize each other. For example simultaneous increases in demand and supply tend to neutralize each other as regards their effect on price. If the increase in demand were the predominant change,

then price would tend to rise in spite of the counteracting influence of the increase in supply. If the increase in supply were predominant, the price would tend to fall. Finally the two increases might exactly neutralize each other and so cause no change in price.

As to the quantity exchanged, the combined effect of increases in both demand and supply must be an increase, and a greater increase than would have been caused by either force acting alone.

The reader should now be able to develop the consequences of other combinations of changes in demand and supply.

Demand and supply the sole determinants. Price changes are caused by changes in demand and supply, and these are the sole determinants. This may at first seem a sweeping assertion. Is not the price of eggs affected by the laying activity of the hens? Does not the rainfall affect the price of wheat? Are not the prices of many articles affected by changes in fashion? Was not the price of liver raised by general dissemination of knowledge as to its beneficial effects in human diet? And are we not told that the "high cost of living" is due to "inflation of the currency"?

Such pertinent questions might be multiplied, but their obvious affirmative answers do not imply that here is a host of additional causes affecting competitive price. All these circumstances, and others, do influence prices, but they do so only indirectly. They act only by first affecting either demand or supply. A drought reduces the wheat harvest; the supply is diminished, and the price of wheat goes up. Knowledge of its qualities increased the demand for liver and so raised its price.

There is no side door through which any influence can get at price; price may be approached only through one of the specified main entrances, remembering that we are here concerned only with the prices of goods subject to conditions of pure competition.

The retail egg market. The reader may find it instructive to compare the behavior of the hypothetical egg market which has served us for illustrative purposes with the records of actual egg markets. The egg market is decidedly seasonal, although the placing of eggs in storage and the withdrawal of eggs from storage in periods of flush and slack production respectively tend to even out the supply during the year.

The interrelation of these factors of production and price are clearly illustrated by the market record which is here presented, statistically in the table on page 121 and graphically in Figure 9.

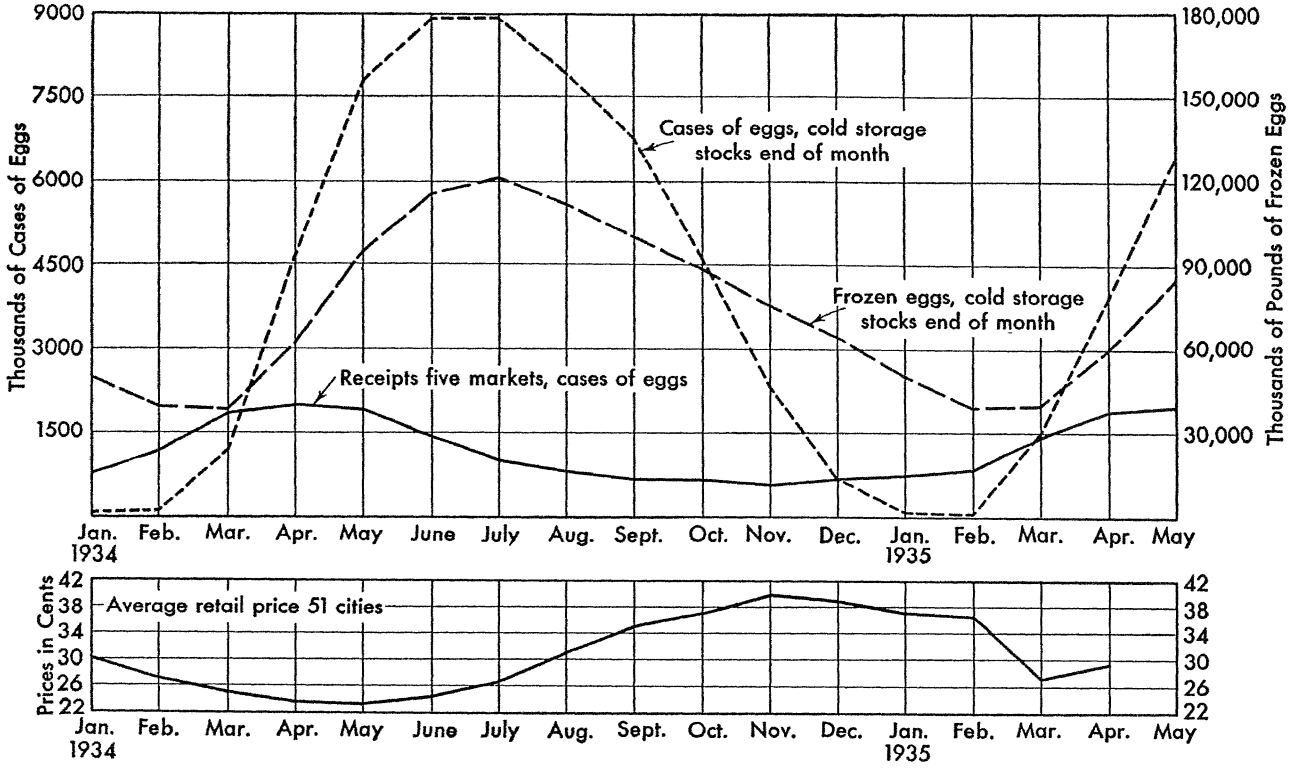


FIG. 9. EGG MARKET STATISTICS, 1934–35 [1]

[1] Figures taken from accompanying table on page 121.

These figures relate to a number of markets combined, but they illustrate what happens in any one market. Receipts of eggs

EGG MARKET STATISTICS[1]

YEAR AND MONTH	RECEIPTS FIVE MARKETS	COLD STORAGE STOCKS END OF MONTH		AVERAGE RETAIL PRICE 51 CITIES
	(Thousands of cases)	Cold Storage (Thousands of cases)	Frozen (Thousands of pounds)	Cents per dozen
1934				
January	808	50	49,910	30.1
February	1,165	90	39,181	26.9
March	1,824	1,208	38,679	24.7
April	2,051	4,640	62,632	23.8
May	1,927	7,819	93,947	23.4
June	1,452	8,965	116,058	24.2
July	1,009	8,961	121,564	26.5
August	828	7,938	111,994	31.6
September	665	6,803	99,951	34.8
October	655	4,633	88,715	36.7
November	588	2,380	76,073	39.7
December	642	648	64,879	38.5
1935				
January	750	39	52,726	37.7
February	858	34	39,413	36.7
March	1,488	1,508	39,516	28.6
April	1,866	3,901	59,313	29.6
May	1,963	6,366	84,741	——

[1] Receipts and cold storage stocks from *Survey of Current Business*, Jan., 1935, p. 42, and July, 1935, p. 44; prices are monthly averages of biweekly quotations in various issues of *Monthly Labor Review* (U. S. Bureau of Labor Statistics).

rise during the first of the year until they reach a high point in April or May. Under the steady influence of increasing supply, egg prices drop off in the spring, reaching a low point at about the same time that receipts of eggs are greatest.

Part of the excess production of the spring and early summer is withdrawn from the market for cold storage. Cold storage holdings reach a minimum about the end of February. They accumulate rapidly during April, May, and June, when prices are low, and reach a peak about the end of July.

As receipts begin to fall off during the late summer, the price of eggs gradually rises, and storage holdings gradually decline. Prices rise during this period and continue to do so until the end of November. During the winter months the price of eggs grad-

ually declines but stays on a relatively high level as compared with the summer months.[1]

The basic law of price change. The conclusions thus far separately arrived at may now be summarized in the following basic generalizations: *When there is pure competition of sellers,* (1) *the direct effect of a change in magnitude of demand tends to be a like change in price;* (2) *the direct effect of a change in the magnitude of supply tends to be an opposite change in price; and* (3) *the direct effect of a change in the magnitude of either demand or supply tends to be a like change in the quantity exchanged.*

EXERCISES

1. Draw up an assumed schedule of the demand for potatoes on a certain day in a small town.
 (*a*) Construct the demand curve corresponding to this demand schedule.
 (*b*) Take any point on the demand curve, and draw perpendiculars from this point to the horizontal and vertical axes. Exactly what does each perpendicular represent?
2. What are likely to be the consequences when potential buyers are informed about the available supply or sellers are informed as to the number of buyers or the prices at which they would be willing to buy?
3. Assume the following demand and supply schedule for fresh eggs, medium size, in a certain city on a certain day:

PRICE	QUANTITY THAT WOULD BE	
	Purchased	Offered for sale
$0.30	1,300 dozen	200 dozen
0.31	1,140 dozen	370 dozen
0.32	1,000 dozen	530 dozen
0.33	880 dozen	670 dozen
0.34	780 dozen	780 dozen
0.35	680 dozen	900 dozen
0.36	600 dozen	1,000 dozen
0.37	530 dozen	1,080 dozen
0.38	475 dozen	1,160 dozen
0.39	410 dozen	1,230 dozen
0.40	350 dozen	1,290 dozen

 (*a*) Plot the corresponding demand and supply curves.
 (*b*) Draw, on the same diagram, a new demand curve indicating an increased demand and a new supply curve indicating a decreased supply.
 (*c*) What is the new price of eggs? Explain the change.
 (*d*) What is now the quantity exchanged? Explain the change.

[1] For another record of the actual egg market, cf. Vanderblue, H. B., *Economic Principles.* Chicago: A. W. Shaw Company, 1927, pp. 110–16.

4. (*a*) Plot again the demand and supply curves corresponding to the schedules of Exercise 3.
 (*b*) Draw, on the same diagram, new curves indicating an increased demand and an increased supply.
 (*c*) What is the new price of eggs? Explain the change.
 (*d*) What is now the quantity exchanged.
5. Using the definitions of demand and supply as adopted in this book, which of the following expressions is not permissible? Explain your answer and draw simple diagrams to illustrate each case.
 (*a*) A cold winter increases the demand for coal.
 (*b*) An increase in price would reduce the demand for coal.
 (*c*) The supply of automobiles exceeds the demand.
 (*d*) Favorable weather conditions increase the supply of wheat.
 (*e*) Demand and supply are equal at the market price.
 (*f*) The greater the demand, the higher the price.
 (*g*) If manufacturers would reduce the prices of electrical appliances, demand would increase.
 (*h*) An armament boom will increase the demand for copper.
 (*i*) An increase in demand causes an increase in supply.
 (*j*) The higher the price, the less the demand.
6. What changes in demand alone or in supply alone or in demand and supply simultaneously would lead to the results indicated below? Draw simple diagrams to illustrate in each case.
 (*a*) An increase in price.
 (*b*) A decline in price.
 (*c*) An increase in quantity exchanged.
 (*d*) A decrease in quanity exchanged.
7. What simultaneous changes in demand and supply would fail to cause (*a*) any change in price? (*b*) any change in quantity exchanged?
8. (*a*) Although the demand for farm wagons has fallen off enormously in the last thirty years, a wagon costs somewhat more today than it did in 1910. What is the explanation? Draw a diagram to indicate what has happened.
 (*b*) "Prices of automobile tires have steadily been reduced, in spite of an enormous increase in demand. This is an illustration of the failure of the law of supply and demand to function." Do you agree? Explain. Draw a diagram to illustrate what has occurred.
9. How would you explain the following price comparisons? Draw simple diagrams to illustrate in each case.
 (*a*) The price of cement was the same in 1930 as in 1900.
 (*b*) The price of a standard grade of lumber was higher in 1930 than in 1900.
 (*c*) The price of gasoline was much lower in 1938 than in 1908.
 (*d*) The price of an ice refrigerator was the same in 1938 as in 1900.
 (*e*) The price of sugar in 1938 was much lower than in 1860.
10. A manufacturer asks for bids to furnish him 100 machines of a particular type. He is prepared to pay any price up to $150 apiece.

(*a*) Draw a diagram illustrating this situation. What price does he have to pay?

(*b*) Draw supply and demand curves to indicate a condition under which there could be no sale.

CHAPTER VIII

Demand

Introduction. In the early days of the science of economics, a certain cynical observer is reported to have said: "If you want to make a first-class economist, catch a parrot and teach him to say 'supply and demand' in response to every question you ask him." Lest our discussion of equilibrium price in the last chapter should seem to offer justification of such an unkind remark, we hasten to observe that that chapter was just the beginning of our inquiry into the forces that determine price. We there assumed the general nature of demand and supply on the basis of common knowledge. Before we really come to grips with the subject of price, however, we shall have to learn more about what is back of each of these market forces. The present chapter is therefore devoted to demand.

Demand, of course, represents the buyer's side of the market. In the preceding chapter we used the broker on the stock exchange as the most convenient example of a buyer. But demand operates in all kinds of competitive markets, including those in which manufacturers, farmers, and traders buy their machinery, supplies, and materials, and in which the people make their everyday purchases of consumers' goods. As our next step, it is most important to learn what determines the consumers' demands; *i.e.*, what lies back of the general principle of demand, which we took for granted in the last chapter. In brief, why does the quantity of any good that buyers would take vary inversely with the price?

Consumers' choices. From the broad social point of view, goods are produced in order that human wants may be satisfied. The consumer finds himself confronted with an apparently infinite variety and quantity of goods more or less capable of satisfying his personal wants. He is invited to make his choice, to take whatever things he wishes and as much of each as he may desire, subject only to the condition that he give money in exchange in accordance with a stated scale of prices.

This is indeed an onerous condition, and to the ordinary person

it immediately imposes a stringent limitation upon his purchases. He abandons at once any idea he might have entertained of possessing a magnificent yacht; he turns away from the tailor shop where a suit would be made to his order for one hundred dollars. He finally makes his purchases—from a limited number of goods and in quantities which are strictly limited.

His purchases are the result of a series of decisions, in the making of which he has had to weigh the relative merits of many competing wants, involving the exercise of more or less careful thought and judgment. It is the sum total of these individual judgments and choices of consumers that makes up the demand for goods.

Although it is true that the total demand of consumers is one of the chief price-determining forces, the fact is that the purchases of the ordinary individual are generally only a minute fraction of the total quantity of any good offered for sale. To the individual it appears that prices of the goods in which he is interested are independent quantities, over which he has no control, and we are justified in making the same assumption for the purposes of our present inquiry.

Everyone is interested, as a consumer, in so disposing his limited stock of money among the unlimited number and quantity of goods offered to his choice that his total enjoyment may be as great as possible. And since it is these choices of consumers which determine the kinds and the quantities of the things that may be sold, every producer is interested in the mental processes by which consumers arrive at their choices.

Marginal utility: An example. Let us consider a family living in a village in which there is only one grocery. Let us suppose that on a certain day the housewife finds the grocer no longer offering eggs at the customary price of thirty cents a dozen but on the contrary disposed to charge each separate customer the very highest price he can get. What price would the housewife pay rather than go without eggs? In the first place, she would certainly take count of her present stock. If the egg box at home were quite empty, she would be willing to pay a high price, the alternative being the sacrifice that the family would suffer from going without eggs. If there were one or two dozen on hand, her top price would be less, although she might still be willing to pay a fairly high price rather than go home empty-handed. If instead she had at home five or six dozen, she would perhaps be willing to pay little,

if any, more than the customary price. Finally we can imagine that the stock of eggs at home is so great that our housewife is not interested in buying any, even at the customary price. Then it would take a bargain price to tempt her to buy.

We can change our supposition by supposing that, with no possibility of purchasing eggs, the housewife were asked to sell a dozen eggs. Clearly the sacrifice suffered through giving up a dozen eggs would depend upon the number of eggs on hand. The storeroom might be so well supplied that the housewife would part with a dozen for thirty cents. On the other hand, were she down to her last dozen, she might think a dollar or two was not too much to compensate her for doing without eggs.

Utility. Utility has previously been defined as that quality of wealth of free persons which enables them to satisfy human wants. Utility for our housewife has to do with the relation of utility to herself, the person whose wants are to be satisfied. From her standpoint, utility is viewed subjectively, rather than objectively, and can without inconsistency be defined as follows: *The utility of anything expresses the satisfaction which a person expects to realize from its use.*

Marginal utility. The housewife and her attitude toward the purchase of eggs brings out the important fact that the utility of anything is not an absolute quantity, but (1) is related to a particular person and (2) depends upon the number of units that that person possesses. As a buyer of eggs, the housewife is concerned with the utility of one unit added to or subtracted from her existing stock.

Most of the purchasing decisions of consumers are decisions to buy more or less. Every man owns a stock, large or small, of shirts. It is not the utility to him of this total stock that determines how much he would pay for another, but rather the increase of utility which he would derive from one more shirt. The same is true of food, house space, entertainment, and most of the many other things that enter the consumer's budget.

The concept of the added utility to a consumer of one more unit of a thing is designated in economics as *marginal utility. The marginal utility to a person of a quantity of a good is the difference between the utility of that quantity and the utility of a quantity one unit smaller.* Thus the marginal utility of five pairs of shoes to the owner is the difference between the utility of five pairs and

that of four. The marginal utility to our housewife of six dozen eggs is the difference between the utility of six dozen and the utility of five dozen.

Properly understood, we could paraphrase this definition as "marginal utility is the utility of the last unit." But we must take care to note that the units are all alike. Once in possession of the user, any unit will render him equal service, afford him equal satisfaction, no matter in what order it may have been acquired. If a difference in utility is occasioned by the addition or subtraction of a particular unit, this change cannot be attributed to any particular characteristic of that unit. Any other unit would produce the same result. The "last unit" does not refer to any particular unit; it is merely the unit that would bring the stock up to its present quantity.

The problem of measurement. Satisfactions, being psychological sensations, are not readily and directly measurable. The utility of a good or the marginal utility of any quantity of it is not susceptible to any absolute measurement; that is, measurement by itself.

When our housewife finally offers thirty cents for one more dozen eggs, it might be presumed that by this act she has measured the marginal utility of eggs to her at that time in units of money. And so she has, but her measurement has not been absolute but relative. Her willingness to pay thirty cents for another dozen is not solely a measure of the satisfaction she expects to derive from these additional eggs. Her decision is also a measure of the utility to her of the thirty cents she will pay for those eggs, and this in turn depends, first, upon her total income, and secondly, upon the utilities at that time of the various other things she might purchase with her thirty cents. It follows therefore that satisfactions cannot be measured absolutely in terms of money.

However, a relative measure of satisfactions to be derived by any particular person at any given time can be obtained, and fortunately this is all that is required for the study of demand. Here utility is important as a cause affecting consumers' decisions as to what and how much they will buy. Every such decision is the result of a choice among two or more possible purchases. If a man decides to buy a pair of shoes for ten dollars, it is because he has decided he would rather have the shoes than anything else he could buy with that sum of money. This implies that he has

balanced the utility of an added pair of shoes against the various other utilities which were equally within his power to acquire.

A perfectly rational and informed consumer would distribute his income so that the last dollar spent on each good and service purchased would yield him an equal amount of satisfaction. This distribution, once achieved, would produce the maximum satisfaction attainable from his income, for to divert a dollar from any good to any other good would involve substituting a lower (marginal) utility for a higher (marginal) utility and would thus result in a net loss of satisfaction.

Every buying decision of a consumer is thus the result of a balancing of marginal utilities against one another. How a consumer will distribute his income among the different objects of expenditure depends upon the marginal utilities of the various goods relative to each other.

To find the marginal utility of good A to a person in terms of good B, we have only to find how much of B we would have to give him to make him just willing to part with one unit of A. A boy with a bag containing a certain number of apples might be willing to trade one apple for seven marbles or for six marbles or for five marbles, but not for four marbles. Thus we can say that the marginal utility to him of apples (under the circumstances of his possessing this particular number of apples) in terms of marbles is five. We may assume also that the marginal utility of marbles in terms of apples is about one-fifth, or at any rate something between one-fourth and one-fifth.[1]

Now if instead of measuring the marginal utility of apples relative to marbles and of marbles relative to apples, we measure the marginal utility of each good relative to all other goods together, we have what influences the ordinary decisions of consumers. To each individual his money represents a certain quantity of purchasing power, which may be used to buy any of the things that normally enter into his consumption. Therefore the purchase of, let us say, a dozen eggs is really a transaction in which the individual exchanges for the eggs a certain amount of purchasing power, which is valuable because of all the other things which

[1] Making our units small enough, we could theoretically reach the precise point of indifference as to whether the exchange would be made or not. At that point the marginal utility of one good would be the exact reciprocal of the other, each marginal utility being measured in terms of the other good.

it could have bought. And the sum of money he pays is the measure of the marginal utility of the given quantity of eggs relative to other things in general.

Putting this in general terms, we can say that the measure of the marginal utility of any given quantity of a good, relative to other things in general, is the amount of money for which a person possessing that quantity would be willing to part with one unit, or the amount of money that he would be wiling to pay for the unit that would bring his stock up to that quantity.

Let us now refer back to the example of our housewife considering the purchase of eggs. Having none at all in the larder, she would be willing, if necessary, to pay a fairly high price for a dozen eggs. Let us assume that under these circumstances she would pay a dollar (if she had to). If she already had a dozen eggs on hand, another dozen would not be worth the sacrifice of a dollar's worth of other goods, but some smaller amount—seventy-seven cents' worth, let us say. Thus, using figures which, though arbitrary as to absolute amounts, are in proper order of magnitude to one another, we may set up a schedule, showing the respective amounts of money which we assume this person would be willing to sacrifice (if she had to) in order to change her stock of eggs by one unit (the dozen) under various numerical circumstances:

Quantity possessed (unit is 1 dozen)	Number of cents which would be sacrificed in order to gain one unit	Number of cents which would just compensate for the loss of one unit
0	100	
1	77	100
2	60	77
3	48	60
4	38	48
5	30	38
6	24	30
7	18	24
8	13	18
9	9	13
10	5	9
11	2	5
12	0	2
13		0

Such a schedule as this implies the possibility that a person may have so much of a commodity that an added unit—a thirteenth dozen of eggs—would afford him no additional satisfaction at all.

Beyond a certain point additional units of the commodity might indeed lessen his total satisfaction by burdening or inconveniencing him. Thus, extending the schedule, we might show negative satisfactions.

This would be true, however, only on condition that there is no possibility of resale. The ordinary housewife is not interested in buying eggs to sell. But if resale were possible, there would doubtless come a point at which the amount that would be paid for additional units would be based on the money to be received from resale rather than on the satisfaction derived from use. From this point the amount that would be paid would be the same, regardless of the number of units possessed. In the foregoing schedule, for example, the last five lines in columns two and three might all read "thirteen."

Of course it is the third column (not the second) in the above schedule which, according to our definition of marginal utility, measures the marginal utility of eggs in terms of money.

Since it will be understood that we are concerned with relative marginal utility rather than absolute marginal utility, we may take the term marginal utility always to mean marginal utility relative to money. We thus arrive at the measure of marginal utility, as follows: *The marginal utility to a person of a certain quantity of a good is measured by the amount of money that would just compensate him for the loss of one unit of the good.*

Marginal utility curve. The marginal utility schedule (see page 130) is easily translated into a graph, as shown in Figure 10 (on page 132). Taking account of the possibility of resale, this curve might flatten out at its lower end, as indicated by the dotted line.

We must remember that a marginal utility curve does not show a historical sequence of uses or utilities; it depicts a series of simultaneous possibilities, not in the sense that all of them can be realized at the same time, but in the sense that any of them may be realized under the same general circumstances. Thus at a given moment, our imaginary housewife may have thirteen dozen eggs, or she may have none at all, or she may have any intermediate number. The marginal utility curve shows the additional satisfaction, in terms of money, she derives from any dozen eggs which increases her total stock to any indicated number within a given range of possibilities.

Having constructed the curve, we pause to consider its interpre-

tation. It is a marginal utility schedule, a person's mental attitude, converted into graphical form. Every significant point on the curve bears a definite relationship both to the horizontal and to the vertical axis. The distance of such a point from the vertical axis, as measured by the horizontal scale, indicates a certain number of dozens of eggs. The height of this point above the horizontal axis, as measured by the vertical scale, indicates the amount of

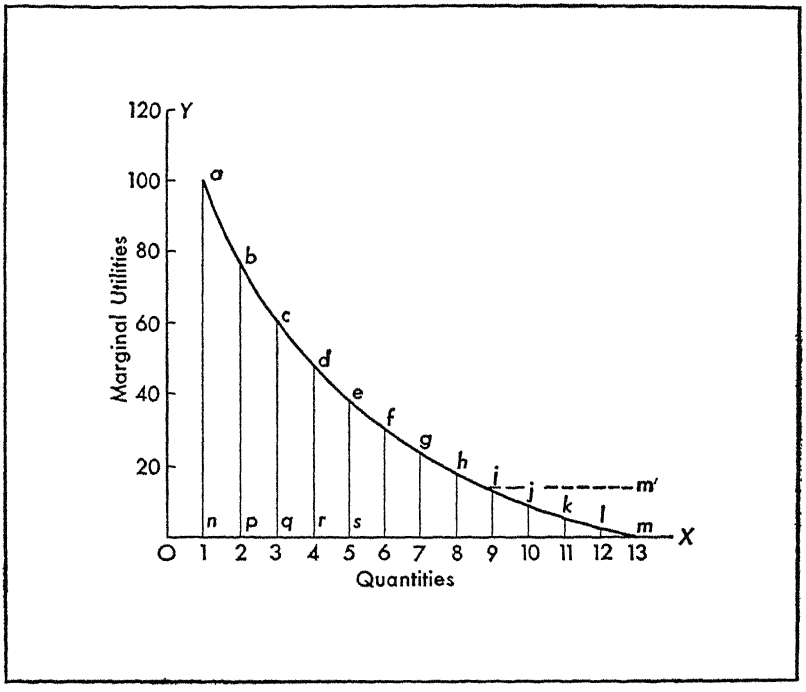


FIG. 10

money our imaginary housewife would be willing to sacrifice to obtain a dozen eggs which increases her total stock to the number indicated.

The law of marginal utility. Once we have a clear idea of the meaning of marginal utility, the economic law of the relation between the quantity of an economic good and its marginal utility becomes obvious. This principle is known as the *law of marginal utility* and may be stated as follows: *At any given time the marginal utility of any good to any person varies inversely with the quantity*

under consideration. A glance at Figure 10 will make the meaning of this perfectly clear.

Marginal utility of durable goods. The principle of marginal utility has been developed thus far with respect to commodities which are rather quickly consumed and of which a person therefore usually possesses several units. One does not usually consider buying more than one lawn-mower, or buying several pianos or dwelling houses. These are durable goods, each furnishing a series of uses for a comparatively long time. One does not ordinarily need more than one unit of any such good. How can the principle of marginal utility apply to such a commodity?

It is a series of satisfactions from the use of a durable commodity that furnishes the scale of diminishing marginal utility. When a person is faced with the alternative of buying one piano or none, he considers the series of uses which a piano would promise him through future years. The marginal utility of the whole series in connection with the price of the piano determines his decision.

Again, having the choice between different qualities of durable goods, a person is often able to fix upon a certain quantity of services in accordance with the principle of marginal utility. In deciding whether to buy a house for $5,000 or $30,000, one is determining how much "house service" he will buy. The marginal utility of a unit of service of the $5,000 house—furnishing not much more than the bare requirements of shelter—will be far greater than that of the much larger series of services—including spaciousness, artistic satisfaction, and the gratification of vanity—which the more costly house offers. Even the durable goods thus conform to the general principle of marginal utility.

Individual demand: Marginal utility and individual demand. The curve *am* in Figure 11 (page 134) is identical with the marginal utility curve in Figure 10 (page 132). It shows the money estimation of the loss which a certain imaginary housewife would suffer by parting with one dozen eggs, assuming her to have any given number on hand, or the price she would be willing to pay for one additional dozen. Every vertical distance (such as *an*) thus measures the price she would be willing to pay for the additional unit which would bring her stock up to the number indicated on the base line, provided, of course, she were held up each time for her highest price.

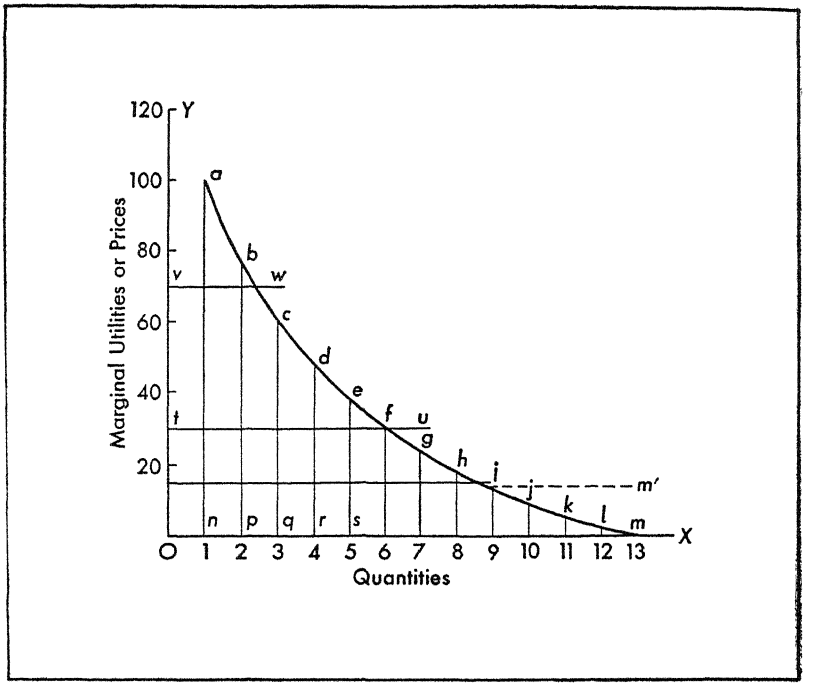


FIG. 11

Now this obviously is not the way eggs are ordinarily bought. There is an established price on the market. No matter how badly the buyer may want the first few units, no matter how great their utility to him, no matter how much he might be willing to give for them if he had to, he does not have to pay more than the market price; and he may ordinarily buy all he desires at that price. The question he has to determine is not what price he will pay, but how many units he will take at the established price.

Suppose our housewife finds that the price of eggs is thirty cents a dozen. How many dozen will she buy? Referring to Figure 11, we conclude that she will certainly buy one; rather than go without one dozen she would be willing to pay a dollar. Similarly, she would buy a second, a third, a fourth, and a fifth, since in each case the sum she would give if she had to is greater than the price. For a sixth dozen she would give thirty cents. She would just get her money's worth by buying this one. A seventh dozen would be

worth only twenty-four cents to her. She would gladly have this one and several more, but not at thirty cents. We conclude that this person would buy six dozen eggs.

The curve in Figure 11, which originally illustrated the marginal utility of eggs under different circumstances, thus serves equally to illustrate the quantities that an individual would buy at various prices. From the point of the *y* axis which measures thirty cents, we draw a horizontal line, *tu*. This line intersects the curve at the point *f*, indicating that at a price of thirty cents six dozen will be bought.

Let us suppose our housewife had found that they were charging seventy cents. She would consider this an exorbitant price; and while she would get enough to supply her most pressing needs, she would buy fewer than if the price were more reasonable, in this case two dozen, as indicated by the horizontal line *vw*. If on the other hand, she had found that eggs were being sold at fifteen cents a dozen—a bargain price—she would have bought a few more; namely, eight dozen.

Of course only one of these price possibilities can be realized at a given time. Suppose the price is actually announced to be thirty cents. This individual will buy six units. This act should not blind us to the fact that consciously or unconsciously there resides in her at the same time both the willingness and the ability to buy eight units at eighteen cents, seven units at twenty-four cents, and other amounts at other prices as shown on the curve. Such ability and willingness constitute an individual's demand for a commodity.

The concept may be defined thus: *An individual's demand for any good (wealth or service) is a schedule of the respective quantities of that good which he is ready to buy at all possible prices.* A curve, such as that in Figure 11, is the graphical representation of this schedule and is known as the individual demand curve. It is identical with the marginal utility curve. Whichever name it takes, the curve is a picture of possible action in response to possible prices. It is controlled, not by the marginal desirability of the specified commodity alone, but by a balance of the desirability of that commodity against the desirability of other commodities and services.

Demand and desire. Demand is not the same as desire. An impecunious student, standing with empty pockets ouside a theatre, may have a keen desire to see the show, but we cannot say that he

has a demand for a ticket. His desire has no effect upon the demand for tickets, no effect upon the numbers that could be sold at various prices. To have a demand for anything one must not only desire it; he must also be able and willing to buy. The only one who contributes to the demand for a certain make of automobile, for example, is the one who is able and willing to buy at least one at a price at which this automobile is likely to be for sale. Though human wants are obviously the origin of the demand for any good, we need to be on our guard against thinking that desire and demand are the same thing.

Significance of this analysis. Hardly anyone would actually take the trouble to make minute calculations such as have been assumed in the foregoing analysis. Nevertheless most of us do have to exercise judgment in the expenditure of money, whether it be only a dollar or ten thousand dollars. Everyone has at least in his subconscious mind some idea of the satisfaction he expects to get from his purchases and can sense to some extent the result of spending a little more on one thing and correspondingly less on others. The given examples are more precise and definite than ordinary human behavior, but such examples are necessary for an insight into the complexities of human problems. A map is far simpler than the country it describes. A globe in the library is smoother than the earth it represents. For the same reason illustrations of economic principles often have to be simplified at the expense of absolute realism.

The preceding analysis illustrates (admittedly in ultra-simple and definitive fashion) the manner in which an individual weighs marginal utilities against each other, adjusting his expenditures when he finds marginal utilities out of balance, seeking marginal utilities proportional to the prices he has to pay—all in the effort to make the most of the limited purchasing power at his command. The thought, the feeling, the unseen reflexes which govern this balancing of utilities, may be known only to the individual himself, and even then very imperfectly understood; but out of it all arises visible action, the purchase of economic goods.

How both parties to a trade can gain. People sometimes find it difficult to believe that both parties to a trade can gain. It has been somewhat plausibly argued that, since in any honest trade the values of the two things exchanged must be the same, neither party can have made a gain; one party can gain only by giving

less value than he receives; that is, by cheating the other party. It was partly this notion which led the churchmen and lawmakers of the Middle Ages to believe that trade should be strictly regulated and customary prices enforced in order that neither party to a trade should gain at the expense of the other.

But unless the marginal utilities a buyer has in mind are born in ignorance and false anticipation, a buyer gains by the opportunity to purchase any commodity so long as the amount obtainable for the money spent more than compensates him for the loss of the money. The seller, on the other hand, has received more than enough money to compensate him for the loss of the goods with which he parted.

Figure 11 gives us an indication of the buyer's gain. Suppose the price is thirty cents. This individual would buy six units. From the shape of his demand curve, which measures the marginal utility to him of various amounts of the good, we can tell that the first unit of the good would compensate him for the loss of one dollar, the second for the loss of seventy-seven cents, and so on. We can be certain that each of the six units which he purchases more than compensates him for the thirty cents which he must spend to acquire it[1] and that if his judgment has been sound he is indubitably better off after the transaction than he was before.

Real people. Of course the ordinary person does not use marginal utility curves or schedules or demand curves or schedules in determining what he is actually going to do; they are, for the economist, ways of sketching the various things an individual may do in response to various possible stimuli.

The closest approach to rational buying is seen when people budget all their expenditures. A careful family budget is an unconscious illustration of the principle of balancing marginal utilities. But many, perhaps most, people do not go even so far as the family budget. Poor people and even the moderately well-to-do often appear to spend their limited stores of money in the most foolish and irrational manner. A family that is without means to provide warm clothing for the coming winter will cheerfully spend ten dollars on a day's outing at the seashore. The radio, with its monthly installment payments, may be eating up the money that would seem much better spent for nourishing food. Clever adver-

[1] Except possibly the marginal unit, which may, as at six units, exactly compensate for the price.

tising and smart salesmanship bank on the careless and irrational way in which people spend their money. All sorts of appeals are used to induce the careless purchaser to buy things whose utility seems absurdly small in comparison with the money expended. To some the appeal of a "bargain" is irresistible, and people come home laden with things for which they have no real need, bought simply because they were cheap; every retail merchant takes advantage of this human characteristic. In any city, on the other hand, it is possible to see people buying things in one store when the identical articles are on sale at a lower price only a block or two away.

Is all buying irrational? From such facts as these we may be tempted to draw the conclusion that there is no such thing as rational judgment in buying, that the principle of balancing utilities has no existence outside the imagination of the theoretical economist, and that all our talk of demand and supply is pure theory without any useful application to the hard facts of the practical world.

But let us not be too hasty in jumping at such conclusions. Much of the conduct of other people which we regard as irrational is not really so but is merely at variance with our own judgment. We readily think we could lay out a family budget for our impecunious neighbor which would materially increase the utility of his purchases. But it is his judgment, such as it is—not ours which decides how he can get the most for his money. The well-fed, well-dressed woman of the "upper class" points with scorn at the "irrational" conduct of the stenographer taking her pathetically inadequate luncheon at the soda fountain in order to save money to buy nylon stockings. But this conduct, whatever else may be said of it, is not irrational. The girl purposely spends more money on clothes, and less on food; this is her way of balancing marginal utilities. Her conduct may be "foolish" according to some other person's judgment. But the economist is investigating how she actually spends her money, not how she "ought" to spend it.

Let us reverse the case. The shopgirl laughs at the wealthy patrons of the Fifth Avenue store who "haven't sense enough" to walk over to Sixth Avenue and buy the same things at lower prices. But it does not follow that the wealthy shopper is irrational. She wants the calm restfulness, the air of luxury and dignity, the deferential personal attention of the exclusive Fifth Avenue shop.

These are part of what she is paying for. She knows she could save money on Sixth Avenue, but the loss of these other advantages would mean a net sacrifice in the utility of her purchases. She also may be "foolish," but that is beside the present point.

Total demand: Individual demand and total demand. Thus far we have considered demand from the standpoint of the individual purchaser, who regards the prices of what he would buy as fixed and not subject to his control. But though the individual's demand for any good has no appreciable effect upon its price, the total demand of all consumers has a profound effect. The total demand is the sum of the demands of all individuals who are contemplating buying the good in question.

For the sake of a simple concrete example, let us assume that in a certain village four families only[1] are interested in the purchase of strawberries, and let us assume their respective demands as shown in the accompanying schedule. Family A is evidently well-to-do; they will take all the strawberries they want, regardless of what the price may be (within the limits of possible prices).

DEMAND FOR STRAWBERRIES, RETAIL MARKET OF X—MAY 1, 1947

Price per quart	QUANTITIES PEOPLE ARE READY TO BUY				
	Family A	Family B	Family C	Family D	Total
55 cents	10 qts.	2 qts.	0 qts.	0 qts.	12 qts.
50 cents	10 qts.	3 qts.	1 qt.	0 qts.	14 qts.
45 cents	10 qts.	4 qts.	2 qts.	0 qts.	16 qts.
40 cents	10 qts.	5 qts.	2 qts.	1 qt.	18 qts.
35 cents	10 qts.	6 qts.	3 qts.	1 qt.	20 qts.
30 cents	10 qts.	6 qts.	4 qts.	2 qts.	22 qts.
25 cents	10 qts.	6 qts.	5 qts.	3 qts.	24 qts.

To the other families the matter is more serious; their purchases will be governed by the price. The schedule in the last column is the total demand for strawberries in this simplified market.

Definition of demand. It is total demand which operates as one of the chief determinants of price. Hereafter the term demand, when not qualified, will be understood to mean the total demand. It may be defined as follows: *The demand for any good is a schedule of the respective quantities of that good which people are ready to buy at all possible prices.*

[1] The assumption of this small number is of course artificial. It serves to simplify the example without affecting its validity as an illustration of the general principle.

To represent better this definition, let us expand our illustration of the strawberry market of X by including all of the hundreds of families which may be interested in the purchase of strawberries, giving a demand schedule in general character like the following.

DEMAND FOR STRAWBERRIES,
RETAIL MARKET OF X—MAY 1, 1947

Prices	Quantities[a]
55 cents	50 quarts
50 cents	100 quarts
45 cents	160 quarts
40 cents	230 quarts
35 cents	310 quarts
30 cents	400 quarts
25 cents	500 quarts

[a] Hereafter the word quantities in any demand schedule indicates quantities which people are ready to buy at various possible prices.

In Figure 12 this schedule is represented in graphical form, the curve *ag* being the corresponding demand curve.

Time and place of the demand schedule. The demand schedule is a series of simultaneous possibilities. The prices are alternative prices; only one of them can turn out to be the actual price. So of the quantities; only one can prove to be the amount actually bought. Which of the various amounts is actually bought we may not know. If we do, we must not forget that at the same time there is an ability and willingness to buy other amounts at other prices.

The demand schedule is neither history nor prophecy. It does not show a series of purchases which have taken place at such and such prices. It does not show how much will be bought at any particular price after the given date or period. Each day will have a demand schedule of its own, which may or may not be the same as this one.

The demand for anything is stated as of a particular instant of time. However, since the consummation of the exchanges will necessarily take time, the quantities in the demand schedule must express what the buyers, at this instant, are prepared to take during some appreciable period of time.

Theoretically conditions of demand may be changing continuously, so that each succeeding instant has its own demand schedule. For most practical purposes, however, it is sufficiently accurate to consider demand with reference to a considerable period, during

which a certain price might hold and a certain quantity be bought. In the market for certain corporation stocks or certain commodities, such as wheat, cotton, etc., where trading is very active, conditions of demand change rapidly, and there is close approach to the theoretical condition of the constantly changing demand schedule.

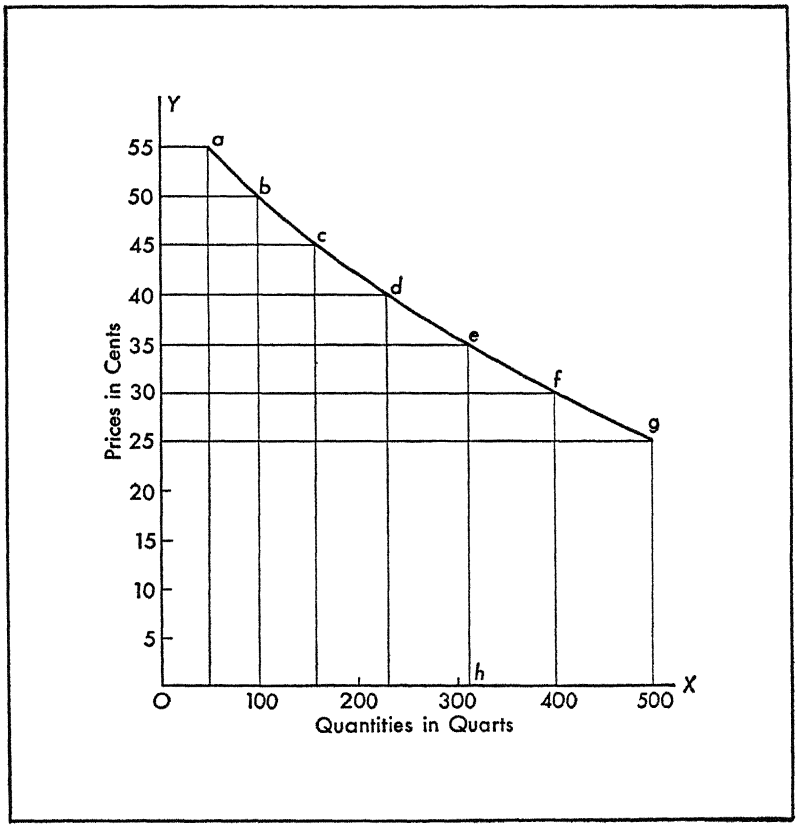


FIG. 12

Of course such a schedule must relate also to a particular place. Market conditions vary, and a schedule showing the quantities which would be bought at various prices in one market will not hold of any other market, except as a coincidence.

Actual demand schedules. The reader hardly needs to be reminded that the exact quantities used in our schedules and graphs

up to this point are not descriptive of any actual market. They are assumed quantities, chosen for purposes of illustration, and not otherwise significant. To ascertain the exact schedule of demand for any commodity at any particular time and place is in fact

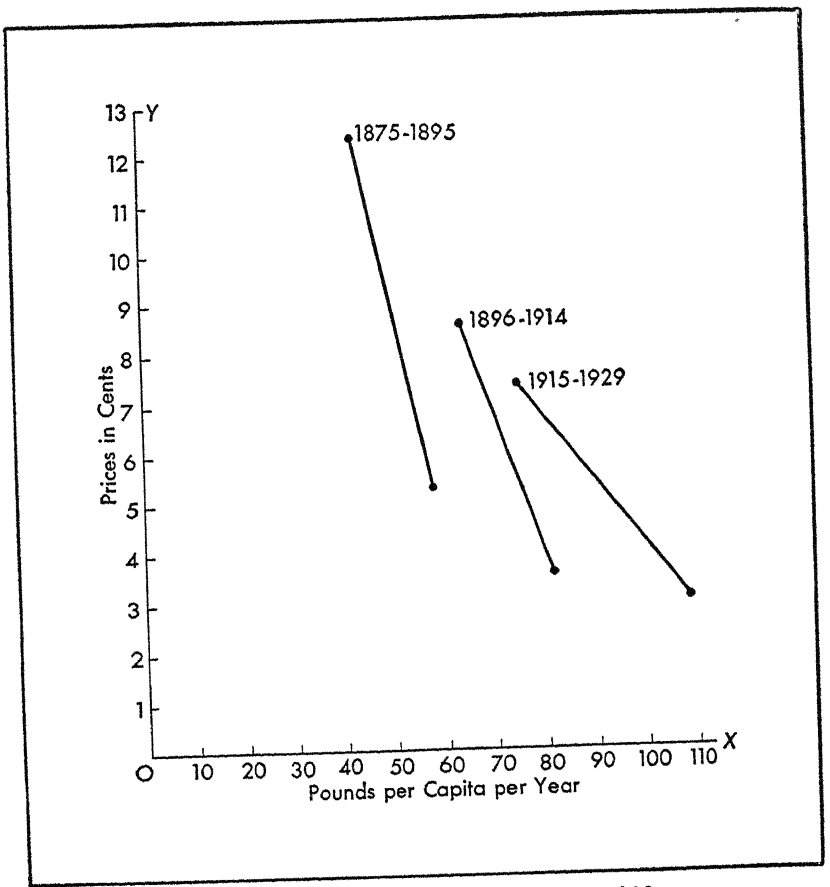


FIG. 13. DEMAND FOR SUGAR, 1875–1929

impossible, depending as it does on the consensus of the independent judgments of many individual buyers, whose judgments, either separately or together, are nowhere recorded.

By means of statistical analysis of the historical record of prices and quantities purchased, economists have, however, succeeded in determining the approximate forms of the demand curves for cer-

tain goods during certain periods in the past. Professor Schultz has thus studied the demands for some ten agricultural products during periods from 1875 to 1929. His results for sugar, for example, are as shown in the graph on page 142 (Figure 13).[1]

The law of demand: General statement. Fortunately it is not necessary for our present purpose—formulation of the law of demand—to have exact schedules or curves of the demands of particular goods. What we require is the general relation between prices and the quantities of goods that people are ready to buy.

We started by recognizing that it is a matter of common knowledge that the higher the price of anything, the less people will wish to buy. This relation between price and quantity has been confirmed by our analysis of consumers' actions based on the principle of marginal utility. It is further confirmed by Professor Schultz's investigation of the demand for sugar. It will be observed that the three curves in Figure 13 have a downward slope going from left to right. So does the marginal utility curve (see Fig. 11, p. 134), which turned out to be also the individual demand curve. And so also the curve illustrating our hypothetical case of the (total) demand for strawberries, as shown in Figure 12, p. 141.

This downward slope clearly represents a condition in which the quantity taken varies inversely with the price—the higher the price, the less the quantity taken. This relationship is in fact what is known as the *law of demand.* It may be stated as follows: *In general, the quantity of any good which people are ready to buy varies inversely with the price of that good.*[2]

Special cases. We have used the phrase "in general" in the above generalization in recognition of certain special cases in which there is a different relation between price and the quantity that people

[1] The collection of the data and the theoretical and statistical analysis required for the determination of actual demand curves are laborious and intricate.. Cf. Schultz, Henry, *The Theory and Measurement of Demand.* Chicago: The University of Chicago Press, 1938, especially pp. 183, 201, 207.

[2] As we shall use the terms, "inversely" and "directly" do not imply that the variation must necessarily be proportional. If *a* increases when *b* increases, or decreases when *b* decreases, we say that *a* varies directly with *b,* even though the change in *a* may not be exactly proportional to the change in *b;* and correspondingly of an inverse variation.

In accordance with strict mathematical terminology it would be more correct to say that the quantity people are able and willing to buy is "a decreasing function" of the price; but this phrase has been found to be awkward and cumbersome in ordinary speech and calculation. The principle as stated above has abundant precedent in good literary usage.

would buy. If the price of diamonds, for example, should fall very low, it is quite probable that sales, instead of being larger or even as large as they are, would actually be smaller. Desire to have diamonds is stimulated to a large extent by the very fact that they are high-priced. If they should descend to low prices at which everybody could buy them, probably few people would want them. At times when people have plenty of money, the same perversity is sometimes illustrated in the sale of other luxuries and novelties. These exceptions we must recognize when we come upon them, but they should not blind us to the existence of the general principle or tendency.

Balancing marginal utilities. Our investigation of marginal utilities has disclosed the foundation of the general principle of demand. It is not entirely a matter of psychology, since demand depends not merely on desire but also on ability to pay. It depends as well upon an individual's resources and on the way in which he wishes to distribute them among alternative uses. An increase in price not only makes a good less attractive in comparison with others; it also makes it physically impossible for people to buy as much as they could and would at lower prices. A decline in price has an opposite effect purely as a matter of arithmetic.

To understand how individuals distribute their incomes among different uses, we must refer back to the balancing of marginal utilities. If the marginal utilities of two goods do not stand in the same ratio as their prices, an individual will have reason to change the way in which he spends his money. If the price of a good rises he will restrict his purchases of it until its marginal utility has increased by an equivalent amount. If the price falls, he will have reason to expand his purchases so that a general balance can be effected. These considerations people may sense only vaguely. They give them no technical names, but their response is unmistakable. The result is the law of demand.

The law of demand, as exemplified in any particular commodity or service, depends then upon a multiple operation of the law of marginal utility. The marginal utilities of any particular good to various individuals vary inversely with the respective quantities they may consider. At the same time, the marginal utilities of other goods are subject to the same principle. As buyers increase their expenditure on one good, its marginal utility for each of them decreases. Spending more on one thing, they have to cut

down their purchases of others, and the marginal utilities of these other things for each of them increase. Thus each and every buyer is subject to restraint by the counter-working of the same psychological principle. This restraint is intensified by a high price and diminished by a low price. Its outward manifestation is expressed by the law of demand.

As a further practical consideration, possibilities of substitution should be noted. If the price of butter rises too high, some people are prepared to use oleomargarine instead. If lamb chops were no more expensive than Hamburg steak, what a change there would be in many a meat platter!

The operation of the law of demand as it relates to any particular commodity or service may indeed be considered essentially in terms of the availability of substitutes. If a small quantity only of a given food can be sold at a high price, the principal reason may be that with such a high price people prefer to buy substitutes. If a large quantity of a given food can be sold at a low price, the chief reason may be that, because of the low price, people are able and willing to buy this particular good in place of many other possible substitutes.

In a broad sense, all goods are substitutes for one another. When an individual is trying to make up his mind whether to spend more money on clothes and less on a summer vacation, he is comparing two kinds of goods which are not similar. Nevertheless they are substitutes in the sense that his desire for one limits his willingness to spend money on the other. However, although these very dissimilar things must be regarded as substitutes in a broad sense, the demand for a particular brand of suits is likely to be affected intimately only by its availability and the price of close substitutes, such as other brands.

Some apparent price anomalies resolved. In the concept of marginal utility and the law of demand lies the key to certain price relations which sometimes appear to be paradoxical. How is it that some of the more useful things in the world are so cheap, while other things that could be given up without any real suffering are so costly? To some the fact that a diamond may be worth a thousand dollars while a loaf of bread is worth ten cents seems strange, inexplicable, or even wrong.

The answer is that the value of anything depends, not upon somebody's judgment of the importance or righteousness of the

want satisfied, but on demand and supply, and the demand side is controlled by considerations of marginal utility. Marginal utility, we have just seen, varies inversely with the quantity. Diamonds are scarce; their marginal utility is therefore high. Bread is very abundant; its marginal utility is correspondingly low. People are willing therefore to pay high prices for diamonds, but very little for a loaf of bread.

Abundance and scarcity. Some of the most useful things in the world are so abundant that they have no marginal utility at all and no one is willing to pay anything for them. Among the prime necessities which man must have in order to sustain life is air; nothing could be more useful. Yet no one will pay for air, simply because it is so abundant that everyone can have all he wants for nothing. Its marginal utility is zero.

To have value, a thing must have not only utility, but a marginal utility greater than zero. The only things that can have positive marginal utility are those which both have utility and are so limited in quantity that there is not enough to satisfy everybody's wants. It is for this reason that some of the most important things in the world, as sunlight, air, sometimes water, are not considered "economic goods." They have no value, are not bought and sold; it is this, rather than any permanent physical characteristic, that rules them out. Water, free in abundance for all in many country districts, is limited in the city; it there has a marginal utility, has value, and is an economic good.

Elasticity of demand: General statement. Every household uses a certain amount of salt. It would be most uncomfortable to go without salt, or even to cut down appreciably on the usual amount consumed; yet there would be little to be gained by increasing the amount used, since everybody already has about all he wants.

If the price of salt should rise, people would still buy nearly as much as usual. The amount expended for salt is a very small part of the household budget, and even at a higher price, the cost would be small, and the increase could be made up by a slight saving on some other item of expenditure. On the other hand, a lower price of salt would not mean any great increase in the quantity bought, for the simple reason that people were already buying about all they wanted. The law of demand holds good of salt, as of other things; that is, at any given time the quantity that

people are able and willing to buy varies inversely with the price. But it varies only slightly.

For a contrasting example, consider a particular brand of gasoline. People may insist on using about their customary quantities of gasoline, not much influenced by small changes in price. But let the price of a particular company's gasoline rise only a little, and many purchasers will quickly turn to other brands. Likewise a moderate decrease in price will increase sales substantially.

These examples call attention to a very important feature of the demand for any particular good; namely, its *elasticity*. From the law of demand we know that the quantity of any good that buyers will take is governed by the price or, we may say, is responsive to the price. *By elasticity of demand, we mean the degree of responsiveness of quantity to price.*

If a small change in price brings more than a proportional change in the quantity taken, the response is strong, and we say the demand is relatively elastic. If on the other hand a small change in price brings less than a proportional change in the quantity taken, the response is slight, and we say the demand is relatively inelastic.

The dividing line between these two conditions is where a small change in price would bring a proportionate (inverse) change in the quantity taken. This condition is commonly called *unit elasticity. By relatively elastic demand we mean that the elasticity is greater than one; relatively inelastic demand means elasticity less than one.*

Graphical illustration. These concepts may be illustrated graphically. When demand is pictured by the usual curve, the degree of elasticity of the demand at any point on the curve is shown by the slope of the curve at that point.

In Figure 14 (page 148), the curve is so constructed that, at point *p*, a small change in the length of the line measuring price is accompanied by an exactly proportional (inverse) change in the length of the line measuring quantity taken. Thus an increase in price from *Oa* to *Ob* would be accompanied by a decrease in the quantity that buyers would take from *Oc* to *Od*. The changes are measured by *ab* and *cd* respectively. The curve is so constructed that, on the numerical scale used, *ab* is 1/10 of *Ob*, and *cd* is 1/10 of *Oc*. We say, therefore, that the changes are inversely proportional, and the demand portrayed is of unit elasticity.

Elastic demand is illustrated by the curve in Figure 14 at the point *q*. Here a change in price from *Og* to *Oh* brings a change in quantity taken from *Oi* to *Ok*. According to the numerical scale here employed, an increase in price of about 6 per cent brings a decrease in quantity of about 40 per cent. This is a more than proportional change in the quantity taken, showing a relatively elastic demand.

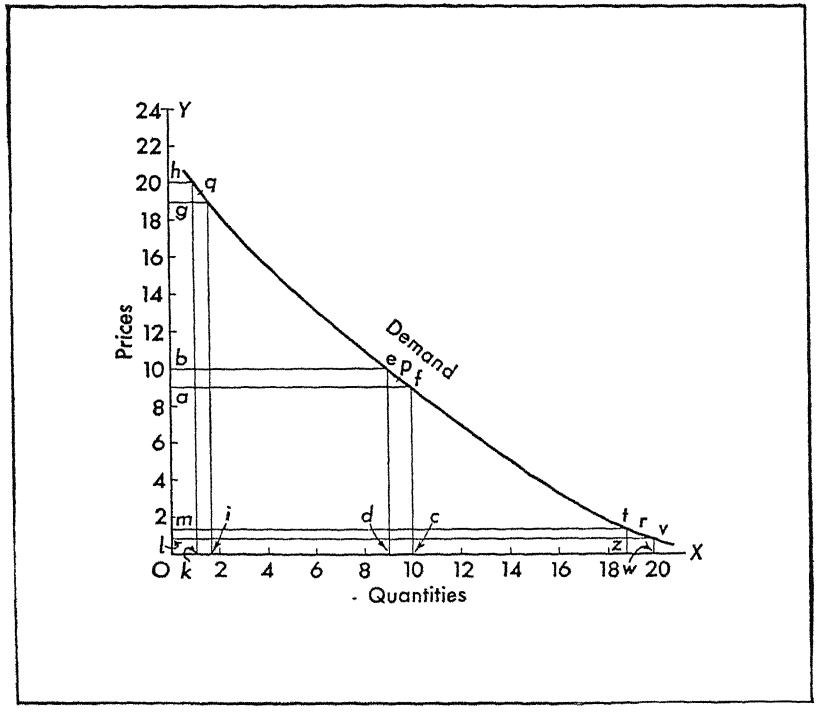


FIG. 14

At the point *r* on the curve, the demand is inelastic. A change in price from *Ol* to *Om* is an increase of about 40 per cent. The corresponding change of quantity taken from *Ow* to *Oz* is a decrease of about 6 per cent. Here we have a less than proportional change in the quantity taken, showing a relatively inelastic demand.

In strict theory statements such as these are precisely true only when demand relates to a mathematical point and the variations in price and quantity are infinitesimal. But businessmen and the consuming public do not think or act in terms of infinitesimals. For

practical purposes we are concerned with appreciable (though small) changes. When, therefore, we speak of demand at a point, we actually have in mind, not a mathematical point, but a short arc of the curve. Our numerical analysis, while thus falling short of strict theoretical precision, is accurate enough for practical purposes.

Elasticity of demand and the amount of money spent. There is an important relationship between price and the quantity that would be taken and the total amount of money that people are ready to spend for any good, depending on the degree of elasticity of the demand.

When the demand is of unit elasticity, the amount of money that people would spend is always the same. This may be demonstrated by the aid of the diagram in Figure 14, p. 148. The amount of money spent is of course the product of the price and the quantity taken. At the price *Oa,* the amount spent is the product of *Oa* and *Oc,* which is represented graphically by the area *Oafc.* At price *Ob,* the amount spent is the product of *Ob* and *Od,* represented by the area *Obed.* These two areas are equal.[1]

When the demand for any good is elastic, the amount of money that people are ready to spend in its purchase varies directly with the quantity bought. When the demand is inelastic, the amount of money that will be spent varies directly with the price.

These propositions regarding elastic and inelastic demand can be readily demonstrated with the aid of the diagram in Figure 14.

Practical examples: Inelastic demand. As a general rule, inelastic demand will be found in the case of simple necessities, since people must have a certain amount, even at the cost of heavy sacrifice of other enjoyments, and since, after their needs are fairly satisfied, they do not much care for an increased amount. Salt is an example of a necessity.

A practical illustration of inelastic demand is exhibited in the following record:

[1] The proof is simple: Let $ab = n\%$ of Ob; and let $cd = n\%$ of Oc.

$$\text{Area } Oafc = Oa \times Oc = \frac{(100 - n)\ Ob \times Oc.}{100}$$

$$\text{Area } Obed = Od \times Ob = \frac{(100 - n)\ Oc \times Ob.}{100}$$

These are identical.

PRICES AND CONSUMPTION OF POTATOES IN THE UNITED STATES[1]

Year	Average wholesale price per bushel for the crop year	Per capita consumption in bushels
1907	$0.61	3.68
1908	.79	3.47
1909	.38	4 34
1910	.48	3.76
1911	1.03	3.26
1912	.49	4.40
1913	.65	3.45

The reader will not have failed to observe that the above table is a record of historical sequence—not a picture of amounts that would be consumed under different assumptions as to price at a given time, as is required by the definitions of demand and elasticity of demand. It is quite safe to infer, however, that the comparatively steady consumption of potatoes in face of a wide fluctuation in price is due to the fact that potatoes are an article of relatively inelastic demand, and that this is true over a wide range of price levels.

The actual demand for sugar, which we have previously noted, furnishes a more scientific example of inelastic demand. See the three demand curves in Figure 13, p. 142. Professor Schultz points out[2] that in the period 1875–95 a drop of 1 cent in price caused an increase of 2.25 pounds in the annual per capita consumption of sugar. In the period 1896–1914, a 1 cent drop in price increased the consumption by 3.34 pounds; in the period 1915–29, the increase was 7.80 pounds. Yet the degree of elasticity was fairly constant through the whole time. In each period, a 1 per cent drop in price increased the annual consumption by about 0.3 to 0.4 per cent. This is a less than proportional reaction and so shows a condition of inelastic demand.

It would be a mistake, however, to conclude that inelasticity of demand is confined to the necessities. Take, for example, tobacco. If by necessity we mean a thing which is required to sustain life or to maintain a decent living in accordance with prevailing stand-

[1] Mills, F. C., *The Behavior of Prices*. New York: National Bureau of Economic Research, Inc., 1927, p. 148.

[2] Schultz, Henry, *The Theory and Measurement of Demand*. Chicago: University of Chicago Press, 1938, p. 229.

ards, then cigars, cigarettes, and other tobacco products certainly are not necessities. Yet few articles of common use show so inelastic a demand. The ordinary man will insist upon having his smoke, no matter what the price. If necessary he will make almost any sacrifice in the other items of his personal or family expenditure. The same is often true of liquor.

Elastic demand. Substitution. The demand is generally elastic for luxuries, things that people could do without. Travel for pleasure is a luxury. The American railroads have learned by experience that a moderate reduction in passenger fares frequently brings a more than proportional increase in travel, indicating a condition of elastic demand.

If luxury is to be regarded as the antithesis of necessity as we have defined the latter term, then the demand for luxuries will as a general rule be elastic. But there will be many and important exceptions; tobacco, for example. Indeed the distinction between necessity and luxury is at best a vague one, and this distinction is far less useful for the scientific analysis of demand and price than the distinctions which rest upon the degree of elasticity of demand.

The most potent cause of elastic demand is the availability of substitutes. The demand for a general line of product (say gasoline) may be fairly inelastic, but the demand for any particular brand of this product is generally very elastic. People may have some preference or other for particular brands, but slight differences in price will swing many of them away from ordinary choices. Suppose the price of a certain company's gasoline should be kept one cent per gallon higher than other brands. How many users of this gasoline would keep on buying it? In general, the demand for a particular brand of almost any product is extremely elastic. The demand for salt in general is very inelastic, as we have observed, but the demand for particular brands of salt is extremely elastic.

This is not only true of particular brands, but also of particular kinds of products. The demand for veal is elastic because people also have beef, pork, mutton, and lamb to choose from. For meat in general there are many substitutes; for example, fish or eggs. These substitutes increase the elasticity of demand for meat.

EXERCISES

1. The marginal utility schedule below reflects the subjective estimate of an individual of the importance to him of successive additional pairs of shoes:

Quantity	Marginal utility
1	$100
2	20
3	12
4	6
5	2
6	1

(*a*) Following the instructions contained in this chapter, construct a diagram to represent this marginal utility schedule.

(*b*) If the particular grade of shoes sells for $10 a pair, how many pairs will this individual find it worth while to possess? Explain.

2. A college student has $3.00 to spend for a ticket to a football game or a dance or a book or a new hat. Explain the basis on which he makes his decision between these alternative forms of expenditure.
3. Assuming that people are frequently influenced by shrewd salesmen, misleading advertising, gossip of friends, etc., to buy goods which give them little real satisfaction, is this inconsistent with the principle of marginal utility?
4. In a period of rapidly changing prices, a person having an annual income which is fixed in monetary terms may be expected to change the nature of his choices among the goods and services available for purchase. Assuming, first, rising prices, and second, falling prices, and applying the marginal utility analyses, show how he would achieve adjustment of his expenditure budget.
5. Draw up an assumed demand schedule for a certain commodity. Plot the corresponding demand curve. Test this demand at two or three widely separated points for degree of elasticity.
6. Explain how each of the following factors tends to affect elasticity of demand:
 (*a*) Necessity or luxury
 (*b*) Presence or absence of substitutes
 (*c*) Habitual consumption
 (*d*) Size of the expenditure for the particular good as compared with one's total expenditure
 (*e*) Advertising warning the buyer to "beware of imitations" or "accept no substitute"
 (*f*) Size of income.
7. In each of the following cases, what condition is implied with respect to elasticity of demand? Explain.
 (*a*) On several occasions during the 1930's the American Telephone and Telegraph Company reduced its toll rates for evening calls.
 (*b*) A small wheat crop may result in a greater income to wheat growers than a large crop.
 (*c*) In 1936 passenger rates on American railroads were reduced from 3 cents to 2 cents per mile, at a time when the railroads were in dire need of additional revenue.
 (*d*) The Brazilian government has frequently resorted to burning cof-

fee or dumping it in the ocean, in an effort to increase the incomes of coffee producers.

8. (*a*) A city calls for bids to furnish it 50 garbage trucks of a particular type. Draw the appropriate demand curve.
 (*b*) The United States Treasury stands ready to purchase all gold brought to it at $35 per ounce. Draw the appropriate demand curve.
 (*c*) What features distinguish these curves from the generality of demand curves discussed, both total and individual? How do you reconcile their shape with the concept of diminishing marginal utility?
9. Does the schedule for Family A on page 139 contradict the principle of diminishing marginal utility? Explain.
10. (*a*) Construct a schedule and a diagram demonstrating the principle, stated in the text, that when demand is inelastic the total amount of money that people are ready to spend varies directly with the price.
 (*b*) Present a corresponding demonstration for the condition of elastic demand.

CHAPTER IX

The Sellers and Their Costs

The different viewpoints of buyers and sellers. The buyers of goods are either consumers, who seek goods to satisfy their wants, or producers,[1] who buy in order, sooner or later, to sell either to consumers or to other producers. The demand of manufacturers and dealers is not the ultimate demand; it is an indirect or reflected demand, arising from the demand of the "ultimate consumers." In any case the goods finally reach consumers, and it is this last exchange which is the fundamental one for the study of the demand side of the forces determining price.

When now we turn to the study of supply, we observe that, with few exceptions, the sellers are not correspondingly parting with goods that might satisfy their wants. There are exceptions, as when a family in reduced circumstances is compelled to sell its household effects, and especially in connection with the sale of nonreproducible goods, such as rare postage stamps, antique furniture, or paintings of old masters. The great bulk of buying and selling however is of goods which are being currently reproduced or are at least capable of reproduction.

Here there is an important difference between the ultimate buyers and the sellers. The purchaser buys a set of golf clubs in order to play the game. The merchant who sells him the clubs is not thereby giving up his opportunity to play golf, and the estimate he places upon the clubs has nothing to do with the pleasure of playing golf. The seller's conduct is not explained by a balancing of marginal utilities.

On the seller's side the motive is profit, or other pecuniary gain. This is true not only of the retail merchant from whom the consumer does most of his buying. It holds also of the wholesaler, the jobber, the manufacturer, the farmer, and all other persons engaged in the various steps of production from the extraction of the raw

[1] The term "producers" includes of course merchants and all others who have a part in the productive process.

material to the delivery of the finished good to the final consumer. This generalization applies also to those who sell their personal services, such as laborers, lawyers, physicians, musicians, and actors.

The seller makes a profit when he sells for more than his goods cost him. He avoids a loss when his selling price just equals the unit cost of his product. Cost of production thus becomes the most important consideration on the seller's side of the market, just as marginal utility is most important on the buyer's side.

Production to order or for a general market. In the early stages of the medieval town economy most of the products of manufacture were "made to order." The blacksmith, the bootmaker, the weaver did not generally begin the making of anything until there had been an order from a buyer and an agreement upon the price. Since no artisan would knowingly undertake to produce and sell an article for a price that would not cover his anticipated costs and a "fair profit," there was here a simple relation between cost and price. Present-day examples of such production are the custom tailor, the custom bootmaker, the house-building contractor, etc.

As a matter of fact a very considerable part of the production of machinery, tools, structural steel, and a host of products that furnish either equipment or raw materials for other factories is made to the customers' orders. The relation between cost and price, however, is not quite so direct and obvious as in the case of the simpler industries cited in the previous paragraph, owing to the complexity of the industrial processes and the consequent difficulty of knowing in advance just what the cost per unit of product is destined to be, and owing also to the different elements that may be included in cost and to the different ways of measuring cost.

The relation between cost and price becomes far more obscure when production is not to order but for a general market. Wheat, cotton, and virtually all agricultural products are obviously in this category. The same is true of most everyday articles which are sold for use to the general public rather than to other producers. Ordinary ready-made clothing, shoes, hats, bread, breakfast foods, and household furniture are a few examples from an almost unlimited list.

In fact this condition is typical of the greater part of modern industry. Such goods are made up in advance of orders from buyers, and the producer takes the chance of being able later to find customers who will take his product at a profitable price. Under

such circumstances we cannot say simply that sellers will not part with their goods except at prices that equal or exceed their costs. The complex relationship between cost and price will appear only after rather searching analysis.

Immediate and remote sales problems. As the first step in our analysis of costs in their relation to price, we need to distinguish between what we may call *immediate* and *remote* sales considerations. The former are those that confront the seller—manufacturer or dealer—with respect to a stock of goods on hand, already produced. Such goods are in a special position, because all of the costs to which their production or acquisition has given rise have already been irrevocably incurred.

The seller's sole interest in them is to realize as large a return as possible now that they have come into his possession. Of course he would like to get back his costs and as much more as possible. But the costs, whatever they are, have already been incurred, and they are generally irrelevant to his decision. In the main this is the situation also of those who sell their personal services, such as lawyers, physicians, and laborers generally. Theirs is an immediate sales problem.

All this is different when the seller is considering the production or acquisition of goods to be sold in the future. With reference to his wheat already harvested, the farmer has to take or leave the going price, regardless of what his wheat cost him. But when he comes to planning next year's crop, he is going to give chief consideration to the prospect of selling his crop for a price at least equal to the cost of raising it. This is what we have called the remote sales problem, and here cost of production becomes the seller's chief consideration. From now on our study of supply will be chiefly directed to cost of production and its influence upon supply.

Supply governed by cost. Businessmen, farmers, and producers generally will not be willing or able to remain in business indefinitely if they sell their products at prices which will not cover their costs. And we must note in this connection that costs include not only the payments made by the entrepreneur to other persons but also interest upon his own capital and an adequate reward for his own labor of management. The owner of capital has the choice of employing it in business or lending it to someone else. In the latter case he receives interest, and he will therefore be

unwilling to employ his capital himself unless he can expect an equal return from his own business. In like manner the entrepreneur generally has before him the two alternatives of running his own business or of hiring out to an employer. Unless the income that accrues to him from his business promises in the long run to be at least as great as the wages or salary he could command, the businessman will ordinarily not continue as an independent producer.

The producer who is unable in the long run to cover his costs, including interest on the capital invested, whether his own or borrowed from others, and wages of management, whether paid to a hired manager or earned by himself, will sooner or later give up and retire from business. The costs of production thus set a minimum below which producers will not in the long run offer to sell their products.

Future cost controls. But the cost of production which thus controls price is not past costs already incurred, but future costs. Suppose the introduction of a new hybrid seed corn has greatly reduced the cost per bushel of raising corn. It will obviously be this new cost upon which the farmers base their plans for growing corn, regardless of the higher cost of stocks on hand. Moreover, since the stocks on hand may now be replenished at lower cost, competition will generally compel the farmers to offer even these stocks at a price related to the new cost of production, no matter what the actual cost may have been.

On the other hand, consider a change which increases the cost of a certain commodity, such as a rise in freight rates or an agreement fixing a higher scale of wages. There will now be a new and higher cost scale governing all future production. Those who have stocks of the commodity on hand, knowing that they cannot be replenished except at greater cost, will find that they can get a price governed by the new cost even for those articles which were produced at the former low cost, and this will control the supply offered by them.

In normal times the prices at which producers decide to offer goods are therefore controlled by future costs of reproduction and little, if at all, by the past costs at which existing stocks were produced. Of course no wise producer closes his eyes to the past or present scale of costs, but it is only as an indication of what future

costs may be expected to be that past or present costs enter into the problem.

Analysis of cost: Fixed and variable costs. Our next task is to obtain a definite idea of the meaning of unit cost of production. For this a careful analysis of cost is required. Let us suppose that a certain manufacturer is producing cotton cloth, using a plant (land, factory, machinery, etc.) which cost $1,000,000. The interest on this capital is, of course, one of the elements in the cost of producing cloth. It is obvious that, this sum having once been invested in the plant, the interest will go on indefinitely, without regard to the amount of cloth produced. Indeed the interest will continue though no cloth at all is produced; it cannot be escaped even by closing down the factory. Even if this particular owner should escape by selling the plant, the purchaser would have to assume the burden. Only by going into bankruptcy can the enterprise escape its contractual interest obligation; and even then the creditors suffer a lump sum loss equal to the present worth of future interest payments.

There is another group of costs whose amount has little dependence upon the immediate number of yards or pounds of cloth produced; this group includes such costs as the salaries of the principal officers and department managers of the factory, the payments for insurance, for lighting and heating the factory, etc. These costs are unlike the interest on capital, however, in that they would generally cease if the plant were closed down permanently.

Finally there are costs which are related closely to the quantity of the product, so that their amount varies with the amount produced. Thus the cost of cotton and other materials, wages, and certain other costs will vary according to the amount of cloth produced by the factory.

Of these three types of production cost, the first and second are called *fixed costs,* because their amount does not vary with the quantity of goods produced. The difference between them is that the interest on the capital invested and permanently sunk in the enterprise must be borne indefinitely (or until bankruptcy), even though the business should be shut down. The other group of fixed costs, on the contrary, would not have to be borne if the business were given up.

The third type of cost is called *variable costs,* because, in contrast to the fixed costs, the amount of the costs varies directly with

the amount of goods produced, though not necessarily in exactly the same proportion.

The classification of production costs may be summed up as follows: (1) fixed costs: (*a*) interest, (*b*) other fixed costs; (2) variable costs.

A numerical example will help us to arrive at the relation between these several groups of costs and the total cost of production per unit of product. Returning to our example of a cotton mill with $1,000,000 of capital invested in plant, let us assume that the annual interest charge, at 4 per cent, is $40,000. Let us assume further that the other fixed costs are $320,000 a year and that the variable costs are 24 cents per pound of cloth produced. What will be the total cost of making a pound of cotton cloth? Evidently it will depend upon the number of pounds produced. The following table shows the results of three separate assumptions as to the output:

COST OF PRODUCING A POUND OF COTTON CLOTH

	First case	Second case	Third case
Amount of product	1,500,000 *pounds*	3,000,000 *pounds*	6,000,000 *pounds*
Costs:			
Interest	$ 40,000	$ 40,000	$ 40,000
Other fixed costs	320,000	320,000	320,000
Variable costs ($0.24 per pound)	360,000	720,000	1,440,000
Total costs	$720,000	$1,080,000	$1,800,000
Average cost per pound	$0.48	$0.36	$0.30

It should be especially noted here that this example does not represent a historical sequence; it sets forth rather the results of three simultaneous possibilities as to the quantity that might be produced. The illustration brings out with clearness the principle that, under these assumptions, the average unit cost of production varies inversely with the amount produced. The reason is simple enough; *i.e.*, since the interest and other fixed costs are constant in amount, the part of such costs assigned to each unit becomes successively less as the number of units increases. There is within limits a general tendency for the unit cost of production of each producer to decrease as he is able to increase the volume of his production.

Economy of fully utilized plant. In this illustration we see unit cost lowered by a "spreading of the overhead." By this we mean

a reduction in average cost through the application of fixed costs to a greater volume of production. This is not the only way in which unit cost may be reduced by increased output. Buying raw materials in larger quantities often makes it possible to procure them at lower prices. Thus an important part of the variable cost embodied in each unit of product may be reduced. Larger production often makes possible greater economy in the use of raw materials and permits the manufacture of by-products which reduce variable unit costs. In similar manner, freight rates and other shipping charges, if included in total cost of a unit of product, may be reduced by an increase in total production. Within certain limits, increase in output also makes possible greater economy in the use of labor.

But there are very definite limits to all these tendencies. Even if 3 million pounds of cloth can be produced more cheaply per pound than 1½ million, and 6 million pounds more cheaply than 3 million, it does not follow that there would be further reduction for 8, 10, 20 million pounds.

Let us return to our numerical illustration, in which it was assumed for simplicity's sake that variable costs were 24 cents a pound. Evidently when 1½ million pounds were made, the plant and the services for which fixed costs were incurred were not being fully utilized; otherwise it would not have been possible to increase the output to 3 million pounds and then to 6 million with the same plant and the same general services.

Sooner or later, if production is continuously increased, the time approaches when plant and general services are fully utilized. There is a general tendency for each producer's unit cost of production to decrease as the volume of his product increases, up to the point where his plant is fully utilized.

Expansion of output and unit cost. If the firm wishes to increase the output beyond this point, it will have to make a choice. Usually it will be possible to increase output further by resorting to various devices which permit still more intensive utilization of the existing plant. If the firm is already running two shifts, a third shift can be added, or the plant may be kept open on Saturdays and Sundays and holidays. Sometimes the actual rates at which operations are performed can be speeded up by increasing substantially the labor force. Still another possibility is that production may be increased by making use of old equipment, less efficient than that ordinarily

employed, which is kept to be used under exceptional circumstances.

If output is increased in this way, it is almost certain to involve a large increase in variable unit costs. If men are hired to work overtime they must usually be paid at a higher rate. As the labor force is increased beyond its normal size, management and supervisory officers are overloaded and function less efficiently. As marginal equipment is called into use, variable costs will be increased disproportionately because of the inferior efficiency of the equipment.

The alternative choice before the businessman who seeks an increase in output is to make additional investment in plant and an increase in the staff of executive officers, etc. This will mean an increase in interest and other fixed costs, and so in the total cost per unit. When the capital investment and the executive staff are increased, the additions will at the start almost certainly be greater than sufficient to care for the increased product which can be immediately sold. Therefore there will again be a situation of unutilized plant, with the possibility of another succession of reductions in unit cost as output is further increased.

The limit to large-scale enterprise. But this will not ordinarily go on indefinitely, since there is in most lines at least a limit to the size of the business unit, beyond which efficiency declines. The laboring force may become too large and unwieldy, wage costs may rise, managerial efficiency may decline through inability to cope with the problems of an overgrown business.

The limit of most efficient size for any particular enterprise depends chiefly upon general conditions in the industry which limit the advantages to be gained from large-scale production. In agriculture, for example, experience has demonstrated the advantage of the small farm. The ordinary farmer soon reaches the point beyond which he does not find it advantageous to enlarge his farm and its equipment. A different situation is illustrated by certain lines of manufacture—such as steel mills, automobiles, oil refining —where there appear to be advantages of large-scale production which are not exhausted until the business unit attains very great size. This situation is still more extreme in the railroad business and certain other public utilities.

The point which especially concerns us here is that each enterprise has a theoretical optimum size, at which under existing conditions it may produce at lowest cost. As demand for its product

increases, the enterprise tends to increase its plant until that size is attained. When all the producers in any industry are in this state, *i.e.,* with enterprises of the size of maximum efficiency and plants generally fully utilized, the limit to the advantages of large-scale production has been reached.

Marginal cost: Definition. When a producer is making a decision regarding a possible policy which would lead to a change in his sales, he must consider two major effects of the proposed step: (1) the change in his revenues and (2) the change in his costs. If the proposed step will reduce costs more than it will reduce revenues or increase costs less than it will increase revenues or reduce costs and increase revenues, then the change will be profitable; otherwise it will not. To arrive at an estimate of the probable effect upon costs, the businessman cannot consider only average costs, with which we have been concerned thus far. He needs the concept of *marginal cost.* We may define *marginal cost* as *the amount per unit by which a small change in output changes total cost.*[1]

The significance of this concept may perhaps be made clearer if we think of marginal cost as the difference in total cost that will be occasioned by a change of one unit in output. This description makes marginal cost exactly parallel with the concept of marginal utility as we have defined it. It is obvious however that, as a practical matter, it is generally impossible to measure the increase in cost that would result if just one more unit were produced, or the saving in cost if one less unit were turned out. For instance, if a shoe factory had been planning to turn out 500,000 pairs of shoes, and we attempted to discover by how much its cost would be increased if it turned out 500,001 pairs, the calculation would be impossible. Of course, the more minutely changes in total cost can be calculated, the more accurate is any measure of marginal cost.[2] In actual practice a fairly rough measure is generally the best that can be obtained.[3]

Marginal cost and average cost. The nature of marginal cost

[1] Since this amount would be calculated by dividing the increase in cost by the number of units added to output, marginal cost is a *rate,* and it is sometimes defined as "the rate at which a small change in output changes total cost."

[2] The pure mathematical concept of marginal cost is based on the assumption of infinitesimal changes in cost and output.

[3] Cost accountants have developed a working concept known as *differential cost,* closely related to the economist's idea of marginal cost.

and its relation to average costs may be given more concrete illustration by returning to the example of the cotton mill which has already served us. Let us revise the assumptions, however, in two respects: in the first place, let us consider this the business unit of most efficient size; and second, let us recognize the probability that variable costs will not be 24 cents a pound throughout.

COST OF PRODUCING A POUND OF COTTON CLOTH

Amount of product	1,450,000 *pounds*	1,500,000 *pounds*
Costs:		
Interest	$ 40,000	$ 40,000
Other fixed costs	320,000	320,000
Variable costs	465,000	480,000
Total cost	$825,000	$840,000
Change in total cost	$15,000	
Change in output	50,000 *pounds*	
Marginal cost	$15,000 ÷ 50,000 = $0.30	
Average of variable costs	480,000 ÷ 1,500,000 = $0.32	
Average of all costs	840,000 ÷ 1,500,000 = $0.56	

In this illustration it is assumed that 50,000 pounds of cloth is the smallest change in the contemplated output of this mill upon which the manager can base a calculation of marginal cost. The effect of a change in production from 1,450,000 pounds to 1,500,000 pounds is to change total costs by $15,000. Dividing this amount by 50,000, we have a fairly accurate measure of marginal cost. The result, 30 cents, is the rate per pound at which total cost changes with an increase in production from 1,450,000 to 1,500,000 pounds, or with a decrease in production from 1,500,000 to 1,450,000 pounds. So 30 cents is said to be the marginal cost of 1,500,000 pounds of cloth.

According to our general definition it would be logical to call 30 cents the marginal cost of either 1,450,000 pounds or 1,500,000 pounds; but to prevent confusion it is generally understood that, when a small change in output takes place through addition, the rate of additional cost is to be taken as the marginal cost of the quantity which includes the added units. Thus the same marginal cost is not attributed to two different quantities; and marginal costs through addition and through subtraction are regarded as the same for any particular quantity.[1]

[1] Compare the similar convention with regard to marginal utility, Chapter VIII.

It is important to note that only in special situations is either the average of variable costs or the average of all costs equal to marginal cost. In the example above, marginal cost (30 cents) is less than the average of variable costs (32 cents), which of course has to be less than the average of all costs (56 cents).

Relation of costs to output. When a businessman undertakes a calculation like this, it makes a good deal of difference where he starts. We have used for our first illustration the problem of increasing the output of a cotton mill from 1,450,000 pounds to 1,500,000 pounds. The three magnitudes at which we arrived—average of variable costs, average of all costs, and marginal cost—as well as their relations to each other, would have worked out differently if we had started with some other output, say 3,000,000 pounds. Without rehearsing the details of calculation under other suppositions, we may present the results of other assumed cases as follows (Case I being the example we have already considered):

SCHEDULE OF UNIT COSTS IN THE PRODUCTION OF COTTON CLOTH

	Possible outputs	Averages of variable costs	Averages of all costs	Marginal costs
Case I	1,500,000 *pounds*	$0.32	$0.56	$0.30
Case II	3,000,000	0.28	0.40	0.24
Case III	6,000,000	0.24	0.30	0.20
Case IV	7,500,000	0.242	0.29	0.29
Case V	9,000,000	0 27	0.31	0.44

It will be observed that this schedule is in harmony with principles which we have already developed, in that the average variable cost, as well as the average of all costs, decreases with an increase in the output, but only up to a certain point, after which each of these averages increases. Marginal cost likewise decreases and then increases. The turning point, however, comes with a smaller output than for either average variable or average total costs; this is because the marginal cost is at first less than the average of all costs and must undergo a substantial increase before its influence will check the decline of average total cost and start it upward.

The reader will of course not fail to notice that we are dealing here, not with a historical sequence, but with a series of simultaneous possibilities, in the sense that, beginning with a given moment, any of the indicated possibilities might be realized at the

expiration of the period necessary for production. The reader will likewise remember that we are dealing with a particular concern, not with a whole industry.

By the same general methods previously used in constructing utility and demand curves, the schedule of costs can be converted into graphical form. Since the table includes, however, three different kinds of unit costs, its graphical representation is not one curve but three, as shown in Figure 15. The curves *cc*, *vv*, and *mm*, respectively, represent various possibilities as to the average of all

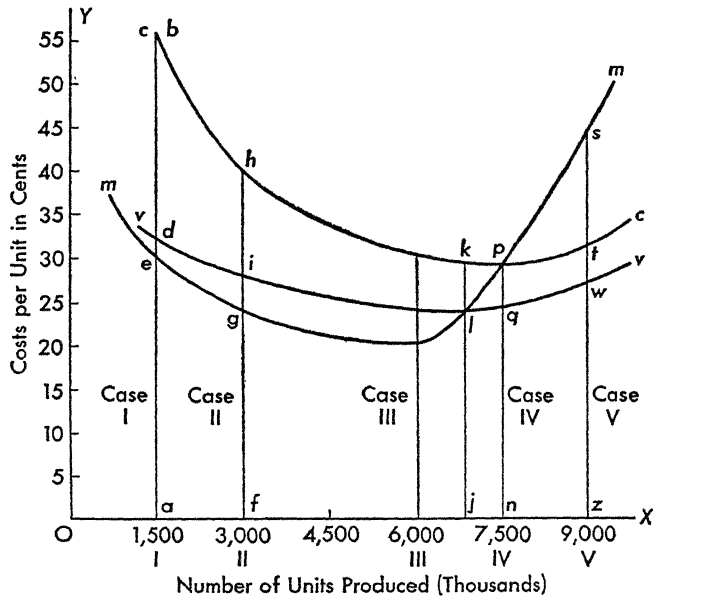


FIG. 15

costs, the average of variable costs, and the marginal cost, under the given circumstances. These curves really show more than is contained in the cost schedule above. Many intermediate cases are understood to be included, and all three curves have been smoothed. If we were to consider changes in interest and other fixed charges,

definite breaks (discontinuities) would have to appear in the curves in Figure 15.

EXERCISES

1. Assume the following conditions of cost for the production of shoes in the X Shoe Company plant:

Output	Fixed cost	Variable cost	Total cost
10,000	$100,000	$ 31,000	$131,000
20,000	100,000	60,000	160,000
30,000	100,000	90,500	190,500
40,000	100,000	123,000	223,000
50,000	100,000	158,000	258,000
60,000	100,000	196,000	296,000
70,000	100,000	239,000	339,000
80,000	100,000	289,000	389,000
90,000	100,000	350,000	450,000
100,000	100,000	429,000	529,000

 (*a*) Prepare a cost schedule showing for each output the average fixed cost, average variable cost, average total cost, and marginal cost of producing that output.
 (*b*) Plot on a sheet of graph paper the corresponding cost curves
 (*c*) Approximately what is the optimum size, or point of maximum efficiency, for this concern?

CHAPTER X

Supply

Marginal revenue. When a businessman is considering the advisability of increasing his sales, he wants to know how this change will affect his total costs and his total revenue. Knowing this, he can judge whether the increase in sales will or will not be profitable. The businessman's cost problem, as we have seen, centers about his marginal cost, the amount per unit by which a small increase in his output changes his total cost. But before he can make his decision, he must have similar knowledge about his revenues. That is, he needs to know his *marginal revenue,* which we define as *the amount per unit by which, under given demand conditions, a small change in volume of sales changes total revenue.*

A purely mathematical concept of marginal revenue would relate to infinitesimal changes in sales, price, and revenue and to theoretical changes as small as one unit in output. Marginal revenue is then the amount added to total revenue by the sale of one additional article. In practice, however, so small a variation can seldom be measured.

As with marginal cost, if the small change in sales takes place through addition, the change per unit in revenue is said to be the marginal revenue of that quantity which includes the added units.

Under pure competition marginal revenue is simply the market price. Each seller's sales are so small a portion of total sales that the amount he offers for sale has no appreciable effect on the price. Each seller therefore considers the market price as fixed. Under these conditions, the total revenues of the seller are increased from each additional unit sold by the amount of the market price.

If for example the price of wheat is $1.25, the farmer knows that, if he sells 100 bushels more, he will add 100 x $1.25, or $125, to his total revenue. The marginal revenue, being by definition the amount added per bushel, is $1.25. In general, any addition to the number of units sold will add to the total revenue a sum equal to the price multiplied by the number of additional units.

Cost of production and seller's policy: Price above average cost. Under any given conditions of marginal revenue and marginal and average costs, what factors determine the quantities which any individual seller will offer at all possible prices? To be specific let us return to our example of the cotton mill and represent its revenue and cost situation by the curves in Figure 16. In addition

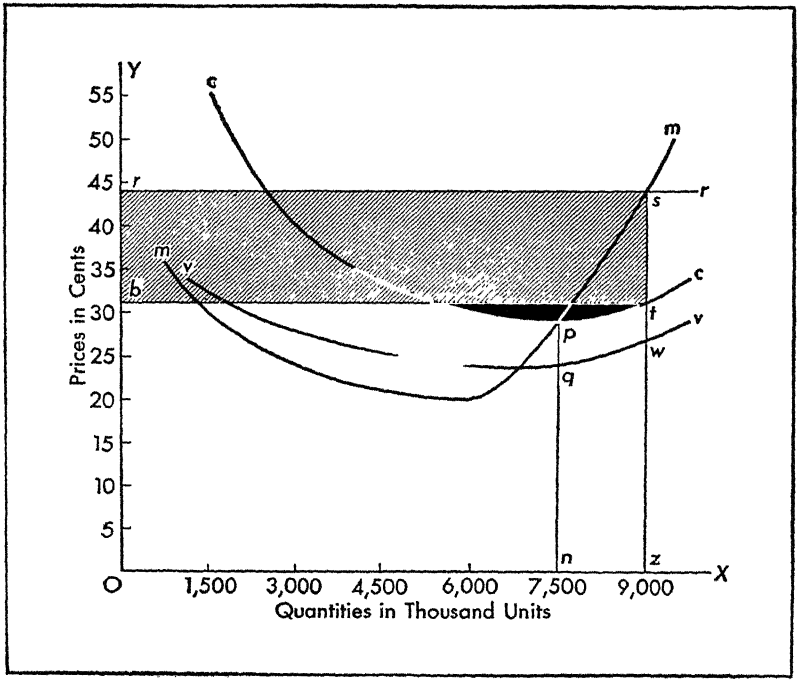


FIG. 16

to the cost curves considered in the preceding chapter there is in the figure the marginal revenue curve *rr*, which under pure competition is, of course, the demand curve for this mill's product and is perfectly elastic.

At the price *zs* (44 cents), the producer will produce and offer for sale *Oz* (9,000,000 units). This price *zs* (44 cents) covers not only *zw*,

the average of variable costs (27 cents), but also *zt*, the average of all costs (31 cents), and permits a clear return of 13 cents a pound. But even more important for our inquiry is the fact that when the producer is making *Oz* (9,000,000 units), his marginal cost, represented also by *zs*, is just equal to his marginal revenue, which is of course the price. Thus the price covers the additional cost incurred by the last increment of production from 8,950,000 to 9,000,000 pounds. It more than covers increases in marginal costs of any lesser amounts as shown by the marginal costs of quantities less than *Oz*.

If the producer were turning out less than *Oz* units, he could increase his return by expanding output. Any increase in output up to *Oz* would add less to his costs than it would to his total receipts. Each additional unit that he sells brings in 44 cents, the ruling price. Each additional unit produced until he reaches *Oz* units increases his total costs by less than 44 cents. Thus it would not pay him to stop short of this output.

On the diagram the net return, or what we may loosely call the total profit, of the producer is represented by the shaded rectangle *rbts*. When selling quantity *Oz* at price *zs* the average profit per unit is represented by *ts*, the difference between the price *zs* and the average cost for this quantity *zt*. Multiplying this average profit *ts* by the quantity sold *Oz* therefore results in the rectangle *rbts*, the total profit of this quantity. This rectangle also represents the difference between the larger rectangle *Ozsr*, representing total receipts, and the rectangle *Oztb*, representing total costs.

Moreover the producer will not be able to sell more than *Oz* units without reducing his total profit. The marginal cost curve, *mm*, is above the marginal revenue curve, *rr*, for every output above 9,000,000 units. If he sells 9,050,000 units rather than 9,000,000 units, the marginal cost curve shows that the additional 50,000 units will increase his total costs at a rate higher than *zs* (44 cents). His total receipts, however, will increase at exactly 44 cents a pound. On any quantity above the 9,000,000 pounds, each additional unit costs more to produce than its sale adds to receipts. It will not pay therefore to produce the extra 50,000 units when the price is 44 cents or for that matter any units beyond the 9,000,000 pounds. If he does, although he may realize a profit from his operations as a whole, his total profit is less than it need be. The producer will then enjoy the maximum profit at a price *zs* (44

cents) when and only when he produces the quantity Oz (9,000,000 units).

By reference to the rising portion of the marginal cost curve, we can determine the amount the producer will sell at any other price, provided the price is greater than np. For any quantity greater than On, marginal cost is higher than average cost. The producer will therefore maximize his profits by increasing his output until marginal cost equals marginal income; i.e. the price.

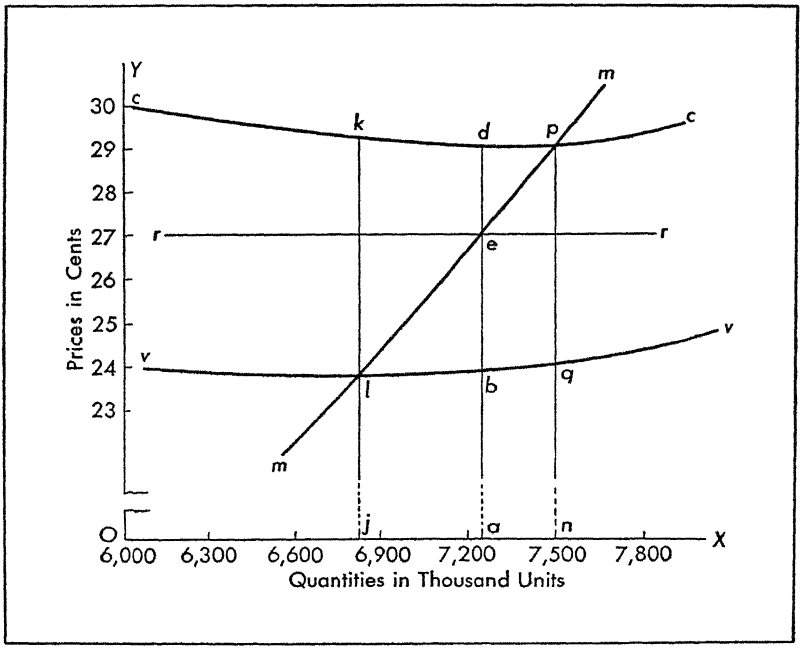


FIG. 17

Price below average cost. But what will be the situation if the price is below the lowest point of the firm's average total cost curve, or below average total cost for any output that might be produced? In Figure 16, the marginal revenue curve, rr, is above the average total cost curve, cc, over a considerable range of quantities. The situation to which our present question refers is represented graphically in Figure 17. This figure may be thought of as an enlargement

of that portion of Figure 16 which includes short segments of the several curves on either side of the line *np* (7,500,000 pounds) where the average total cost curve *cc* is at its lowest level, but with the significant difference that the marginal revenue curve *rr*, representing market price, is below the average total cost curve *cc* for every quantity and specifically below *p*, the lowest point of that curve. There is no output for the firm at which the price is high enough to cover the average total cost of producing that output.

Figure 17 illustrates the first of two possible general situations where price is below the minimum average total unit cost—the situation where price is below that minimum but at the same time above the average variable costs, shown in the figure by the curve *vv*. In this situation the firm is bound to lose money. But since the price is higher than the lowest point on the average variable cost curve, there are certain outputs at which the firm can at least recover its variable costs and have something left with which to meet its fixed costs.

This, though still leaving the firm "in the red," is better than closing the factory, since then its fixed charges would still go on, with no income at all to pay them. Under such conditions therefore it will still be to the firm's advantage to produce such quantity as will make its marginal cost equal to the price. Though a profit is not obtainable, though a loss must be borne, the producer at any rate will thus minimize his loss—for the time being.

Finally, we shall have to consider the second general situation in which the price is below the lowest point of the firm's variable cost curve, *jl* (23.8 cents). Then the cotton manufacturer is unable to recover even his variable costs. No matter what quantity he chooses to produce, he will be continually losing all his fixed costs and something more. Even so, a producer may sometimes continue to operate—temporarily. Anticipating better conditions in the future, he may choose to bear a present loss to keep his plant from deteriorating or to hold together his organization and an experienced labor force. Ordinarily an enterprise that is losing more than its fixed charges will very soon find it best to close down.

Short run and long run. There is thus an important distinction to be made between the costs which a producer must cover *in the short run* and those which he must cover *in the long run*. This distinction is not dependent on a certain interval of calendar time but on technical conditions of the industry in question. The long

run means a period of time sufficiently long to permit, when the industry is profitable, existing firms to expand their plants or new firms to enter the industry with newly constructed plants, or to permit, when the industry is incurring losses, firms already in operation to close down their plants and withdraw from the industry.

The short run means a period sufficiently short to prevent new firms from entering the industry and old firms from expanding if the industry is profitable, and to permit the continued operation of existing firms even though their operations are unprofitable. During the short-run period, a firm will continue production, as we have seen, so long as its variable costs are covered, or so long as the loss from operating the enterprise is less than the loss that would be incurred by shutting down.

Conclusion. From this study three propositions concerning output and price have emerged.

(1) So long as the firm continues in operation under pure competition, it will seek to produce that output for which marginal cost is just equal to marginal revenue, which is the price. This output will maximize its profits when price exceeds average total cost and will minimize its losses when price is below average total cost.

(2) A firm will not continue production in the long run unless it receives a price that is sufficient to cover average total costs.

(3) A firm will continue operations in the short run so long as the price exceeds average variable costs, and it may continue to operate for a short time even if price falls slightly below average variable costs.

These three propositions describe the way in which the firm's cost schedules will determine the amount that it will produce, always assuming that it can sell any amount which it chooses at the going market price.

Individual supply. From the marginal cost curve of any individual firm we can determine the quantities of its product that it will produce and offer for sale at all possible prices both in the short run and in the long run.

As long as marginal cost decreases as the output increases, the firm has certain factors of production which are not being completely utilized. By increasing the variable factors of production it can expand its output and thereby reduce its unit costs. Since any such firm seeks to produce and sell the quantity that will maximize

its profits or minimize its losses, it will continue to expand its output so long as each increase brings a lower marginal cost and therefore a lower average cost. It follows that the firm does not attain stability and arrive at a normal sales policy until it reaches that output where its marginal costs begin to rise.

That portion of the marginal cost curve, therefore, which slopes upward for outputs larger than that represented by its intersection with the average variable cost curve represents graphically the quantity which a firm will offer at each possible price. For our cotton mill this portion of its marginal cost curve showed how many pounds of cloth that mill would be willing to sell at any of the several possible prices. If the price should be below its average variable cost, the mill would soon close down, even in the short run. If the price is between the average variable cost and the average total cost, the marginal cost curve shows what quantities the firm would offer in the short run to minimize its losses. Finally, if the range of prices lies above the average total costs, the marginal cost curve shows the quantities which would be offered to maximize the firm's profits both in the short and in the long run.

Since the marginal cost curve does, therefore, portray a schedule of respective quantities which an individual firm will produce and offer for sale at various prices, it forms the basis of that firm's supply curve. *An individual's supply of any good is a schedule of the respective quantities of that good which he is ready to offer for sale at all possible prices.*

Total supply. Individual supply underlies the concept of total supply, which is of primary importance as one of the factors determining price. When the word "supply" is used without qualification, it will refer to total supply. *The supply of any good is a schedule of the respective quantities of that good which people are ready to offer for sale at all possible prices.* This definition evidently corresponds precisely to the definition of demand.

So close indeed is the analogy between demand and supply that only a few words at this point will suffice to set forth the formal character of the latter. Total supply is the sum of a number of individual supplies. The phrase "ready to offer for sale" implies both ability and willingness. Supply, like demand, relates to a particular place and a particular time; in strictest theory to an instant of time, though for most purposes the time is an appreciable period—a few minutes, an hour, a day, a week, etc., during which the respective

quantities would presumably be offered if various assumed prices should prevail. A supply schedule, like a demand schedule, indicates a series of simultaneous possibilities—on the other side of the market, however.

Supply and stock. Much discussion of the various problems of price goes astray through confusing supply with the stock of goods on hand. For example, what is popularly known as the "visible supply" of various grains in the United States is reported monthly by the United States Census Bureau. These are important quantities, and they have a real bearing upon supply, but they are not the supply of wheat or corn in any market.

The stock of wheat of various grades in the elevators in the neighborhood of Chicago is not the supply of wheat on the Chicago Board of Trade. The supply is the schedule of the respective numbers of bushels of wheat that people are not only able but willing to offer for sale at various prices at a particular time. No one of these quantities will be the same as the quantity of wheat in the Chicago elevators.

In fact, while there is a relation between the quantity of goods on hand and the supply, this relation will often appear not to be very close. The quantity of wheat offered for sale at Chicago at some particular price may be far less than the wheat on hand in the elevators. Much of that wheat the holders may not be willing to sell at that price. On the other hand, the quantity offered might be more than the total stock in the elevators. The Chicago market may be drawing upon wheat stored in other places.

In fact the quantity offered might conceivably be greater than the total stock in existence. This surprising state of affairs occurs now and then in speculative markets. On May 9, 1901, on the New York Stock Exchange, it was suddenly discovered that sellers had contracted to deliver more shares of the stock of the Northern Pacific Railway Company than could possibly be bought or borrowed by them. A disastrous crisis was the result. Of course the people were offering shares which they did not then possess; they hoped to acquire them in time to deliver. But that does not alter the fact that so many shares were actually offered.

In 1917, during the First World War, there occurred a "corner" in wheat on the Chicago Board of Trade. At one time the price was $3.25 a bushel, and it was stated by a well-known authority that "we are getting into a serious situation. Everybody is long, and

there is not available the wheat to deliver to those who have bought it."[1]

Relation of quantity to price: Immediate sales problem. As we have seen, the normal relation between price and quantity offered by an individual seller is that the quantity varies directly with the price. It might appear that the combination of the several individual supplies would lead to a similar relationship for the whole market. This is indeed the result when we are directing our attention to what we have called the "immediate sales problem." If we are thinking only of the quantities that sellers could be induced to part with at once from their existing stocks, it is true, with very few exceptions, that the higher the price the greater will be the quantity offered by sellers.

Remote sales problem: Short run. But when we turn our attention to the remote sales problem, the matter is not quite so simple. Here we are interested in the quantities that will be offered for sale by producers taking into consideration their future costs of production. This problem may conveniently be divided into two subdivisions.

First, how much will be supplied at a given price if production must come entirely from the firms already in operation? It is clear that under this assumption the quantity produced and offered for sale will vary directly with the price. Whether the individual firms are operating at a profit or at a loss, a higher price permits each to increase its output at the expense of higher marginal cost, while a lower price compels each to reduce his product in order to equate his marginal cost with the price. We thus have the normal upward-sloping supply curve.

Long run. Second, what is the relation between price and the quantity offered for sale in the long run: *i.e.,* when time is allowed for firms to enter or to leave the industry? In other words, what is the shape of the long-run supply curve? Here we must consider two sets of conditions as to costs in the industry as a whole, which we call respectively *increasing costs* and *constant costs.*

Increasing costs. Let us start with a condition of equilibrium in which production is in the hands of certain firms who all have their marginal costs equated with the price. We may assume further that there are other firms that might enter the industry, but they could produce only at average unit costs that would be higher than

[1] Patten, J. A., "In the Wheat Pit," *Saturday Evening Post,* Nov. 5, 1927, p. 177.

the present price. In a free economy there will always be such potential producers. It is because of their higher costs of production that, in the present state of equilibrium, they are not already in the market.

Now what would be the consequence of a lower price? Evidently there would be even less inducement for any of the potential producers to enter the market. On the other hand, each of the present producers would reduce the amount of his production, and some of them—the least efficient or marginal ones—would go out of business. The amount offered for sale is thus lower than at the original price.

Consider next the consequences of a higher price. Each of the existing producers would increase his product until his marginal cost was equal to the higher price. And if the price were high enough, some of the potential producers would enter and offer for sale such quantities of the product as would equate their marginal costs with the price. We thus have a situation with an upward-sloping supply curve, the number of competitors and the amount produced by each varying directly with the price.

This state of increasing costs is generally brought about by scarcity of one or more of the factors of production which all firms in the industry require. This condition appears most clearly in the case of agricultural production, where land is the scarce factor. After the early stage of abundant fertile land, which has long since been passed in every well-settled community, a large number of comparatively small farms has generally proved to be the organization of maximum efficiency, and the farms normally have their plants fully utilized. Increasing the product of the existing farms means that more laborers must be employed per acre, more capital must be invested in the form of buildings, farm machinery, tools, drainage systems, etc., more must be spent on fertilizers, and so on. All this means an increased cost per unit of product.

On the other hand, there is only a limited amount of idle land available, and this land is of inferior quality to that now used; otherwise it would be already in use. New farms, because of the inferiority of the lands, will from the beginning have to be operated at higher costs than prevail on the existing farms. Since either means of obtaining an increase in agricultural products thus involves increased unit cost, both will be resorted to, and the increase will come partly from the cultivation of additional inferior lands

and partly from more intensive cultivation of the lands already in use.

Similarly in mining, the output may be increased only through resorting to inferior mines and through working existing mines more intensively by going to greater depths, taking out lower grade ores, installing more expensive machinery, and in other ways adding to the cost for capital and labor more than in proportion to the gain in output. Similar conditions appear in lumbering, fishing, and all the extractive industries by which the raw products of nature are obtained. All these industries are thus normally operated, except in their remote earliest stages, under what is known as the *law of diminishing returns.*[1]

Constant costs. A different supply situation is found in certain lines of industry in which, after a certain price has been attained, it is normally easy for new firms to enter the market with costs no higher, or only a shade higher, than the prevailing price.

For an industry to find itself in this situation two essential conditions must be satisfied: (1) The firms composing the industry must have similar cost curves, and it must be possible for new firms, if they enter the industry, to produce under substantially the same conditions of cost. By similarity of cost we mean, more specifically, that the minimum points on the average cost curves of the various firms must be at about the same level. Usually in such an industry the plant of most efficient size is comparatively small. If the most efficiently sized plant is very large, the industry is almost certain not to be operating under conditions of pure competition, and even if it is, it becomes difficult and hazardous for new enterprises to enter the field.

(2) There must not be heavy reliance upon a scarce factor of production. If there is, the entrance of new firms will raise the general level of costs for all producers so that it will be impossible to expand output except under the stimulus of rising prices.

Under these conditions, supply will expand and contract, not by changes in the output of existing firms in response to price, but by the entry of new firms into the industry or withdrawal of existing firms from the industry. Each producer regards the prevailing price as fixed and will offer only the quantity which equates his marginal cost to the price. Since under pure competition the

[1] For a more elaborate explanation of the principle of diminishing returns, see Chapter XX.

individual firm can increase its output only at increasing costs, no firm already operating will increase its output unless the price rises. But if the price rises, new firms may enter the industry with approximately the same costs as the firms already in the industry. Thus, a small rise in price induces a large expansion of output through the establishment of new firms.

Conversely, a small decrease in price will lead to a large contraction in output. Since firms already in the industry are assumed to have similar cost curves, competition will have driven the price down to the point where it just covers the minimum average cost for each of these firms. It will require therefore only a slight drop in price to render the operations of many of the firms unprofitable and thus, in the long run, to drive them out of the industry entirely.

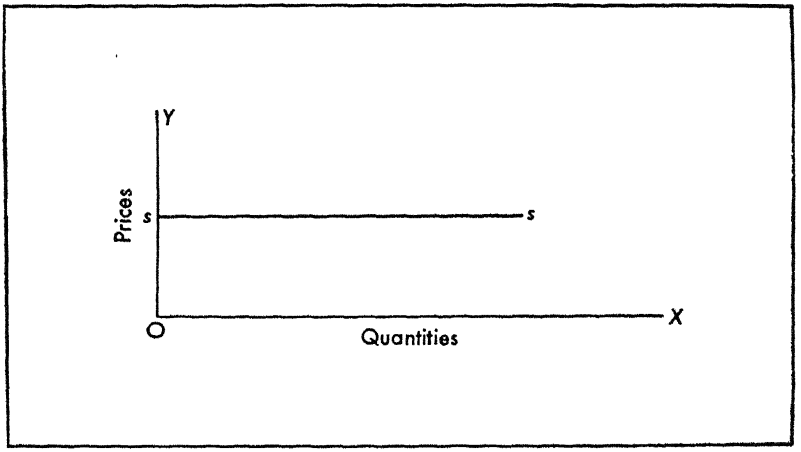


FIG. 18

In this, as in the preceding case, we assume that new firms will be attracted into any purely competitive industry as soon as an opportunity for profit appears there. Whenever the cost of all existing producers and of any new ones that may enter the industry are approximately the same, the price cannot fall appreciably without heavily reducing the output of the industry; nor can it rise appreciably without stimulating a great expansion of output. Each enterprise will have an upward-sloping individual supply curve, but the output of the industry as a whole can be enlarged without increase in the general level of costs. In place of a normal ascending

supply curve we have a horizontal supply curve (illustrated by the line *ss* in Figure 18).

It may at first thought seem paradoxical that, since the total supply is the sum of the several individual supplies, and each individual firm has an upward-sloping supply curve, we can derive a horizontal total supply curve. The paradox is resolved by recalling that, for each individual firm, only one point on its marginal cost curve is relevant for the derivation of a long-run supply curve. This point *p* (see Figure 19) is the minimum point of the average cost curve for a plant of the most efficient size. The height of this

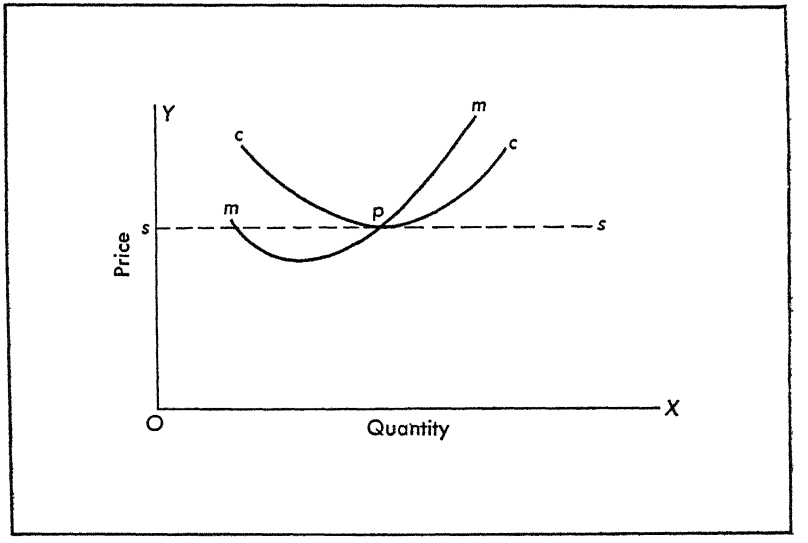


FIG. 19

point above the *x* axis determines the level of the horizontal supply curve *ss* for the industry. Other points on the firms' marginal cost curves are of only short-run significance. For if the price were to rise temporarily above *Os,* new firms would, in time, enter the market and supply as much output as was needed to drive the price down to *Os* again; and conversely if the price fell below *Os,* firms already in the industry would suffer losses, and a sufficient number of these firms would eventually be forced to discontinue operations to reduce supply and to bring price back again to *Os.*

A horizontal supply curve represents for the long run a condition

of perfect elasticity of supply. The constant cost industries are characterized by complete elasticity of supply in the long run, since, when in equilibrium, a slight increase in prospective prices would result in a large increase of product and a slight decrease in price would reduce drastically the quantities offered.

The law of supply. From the foregoing analysis the following generalization, which we may call *the law of supply,* is derived. *In a given market at a given time, the quantity of any good which people are ready to offer for sale varies directly with the price.* It should be noted, however, that at the limit of absolute elasticity of long-run supply a substantial change in price may cause the quantity either to drop to zero or to increase without limit.

Long run and short run again. Of course the practical interests of businessmen and of the consumers generally are more concerned with long-run supply conditions than with the sellers' short-run problems. And the discussion in this chapter has applied primarily to the long run. If, for example, we are studying the variations in the output of cotton cloth that occur within a single year in response to variations in price, it is obviously not realistic to assume that large numbers of new firms will within that period be attracted into the industry by an increasing price, or that existing firms will be driven out of production by a decline in price. For short-run considerations we may generally consider the number of enterprises as fixed, giving us a total supply curve based on the individual supply curves of all existing firms. It is only if we are concerned with long-run tendencies that we may legitimately assume that high profits will attract new producers and that very low profits will drive existing firms out of business.

Though this point is fairly clear in its application to all sorts of production, its significance becomes especially pronounced when we are considering industrial rather than agricultural enterprise. Most manufacturing production is carried on with highly durable equipment, the purchase of which has given rise, as we have seen, to fixed costs. Price might fall to a level well below the minimum average costs of many of the producers in the industry without, in the short run, causing them to cease production. Thus a decline in price will lead to a much smaller reduction of output in the short run than in the long run, just as a rise in price will not have its full effect for a considerable period of time. In our discussion of increasing costs and constant costs, we have found that the shape

of the total supply curve will depend, among other things, upon the ease with which new firms can enter the field. Clearly, therefore, these supply curves refer to the long run rather than the short run.

Elasticity of supply: General statement. There is elasticity of supply of different goods, analogous to the elasticity of demand. Here also, by elasticity we mean the degree of responsiveness of quantity to price.

If a small change in price brings a more than proportional change in the quantity offered for sale, the response is strong, and we say the supply is relatively elastic. A 1 per cent increase in price, for example, would bring an increase of more than 1 per cent in the quantity offered.

If, on the other hand, a small change in price would bring less than a proportional change in the quantity offered, the response is slight, and we say the supply is relatively inelastic. A 1 per cent increase in price, for example, would bring an increase of less than 1 per cent in the quantity offered.

The dividing line between these two conditions is where a small change in price would bring a proportionate change in the quantity offered. A 1 per cent increase in price, for example, would be accompanied by a 1 per cent increase in the quantity offered. This condition is commonly called unit elasticity. By relatively elastic supply we mean that the elasticity is greater than 1; relatively inelastic supply means elasticity less than 1.

Influences affecting elasticity. Elasticity of supply depends chiefly on the relative ease with which factors of production or business enterprises can be transferred into or out of the particular line of production in question. When this is relatively easy, the product of the industry is subject to elastic supply.

The chief characteristics of such an industry may be stated negatively as: (1) no heavy reliance on a scarce factor of production, such as a large area or a particular location of land, or a highly specialized type of skilled labor, or very costly specialized machinery; and (2) no requirement of a very large capital investment or a large and elaborate business organization. We might cite the custom tailors, the house painters, the brick manufacturers, as obvious examples.

When the foregoing conditions are met, it is easy for enterprises to move into and out of the line in question. A small increase in the price of the product will meet a quick response in the form of

large increase in the quantity offered for sale. A small decline in price will, by the same taken, cause a relatively large decline in the quantity offered. The product is one of elastic supply.

Inelastic supply is found when the before-named conditions are reversed, and it is, in consequence, not a simple matter for new firms to enter the industry or for those already engaged to withdraw and go into other lines. Agriculture and mining are limited by the land requirement; steel-making, by the requirement of large capital investment and elaborate organization; railroading, by both land and fixed capital requirements; and there are certain lines of production in which highly skilled specialized labor is the scarce factor. Under these conditions, a considerable change in price may fail to cause such a shift in the industry as to produce a proportional change in the quantity of the product offered for sale. The supply is therefore inelastic.

The limits of elasticity. The absolute limit of elastic supply is reached when sellers are ready to offer any quantity at a given price, but nothing at all, or only a very much smaller quantity, at any lower price. In other words, the price is the same, irrespective of the quantity. The graphical representation of this situation is a straight horizontal supply curve. This condition is characteristic of industries operating at constant cost, as we have just seen.

The absolute limit of inelasticity of supply is reached when changes in the price have no effect whatever upon the quantity offered for sale. Such supply is represented graphically by a vertical straight line. The simplest illustration is the ordinary auction, in which a definite quantity of some commodity is put up for sale. Exactly this quantity will be sold, whatever the price. It is scarcely conceivable that this extreme could be reached in connection with pure competition of many sellers. Since all are subject to rising marginal costs, increase in price would be sure to bring forth increased product from some, if not all, producers.

Decreasing-cost industries. There are certain industries so conditioned that, as the output increases through the addition of new firms, the average costs of the several firms will decline. The following are probably the most important factors which may contribute to this result.

First, each firm may be enabled to concentrate on a narrower range of products or processes, which can then be conducted on a larger scale and at a lower cost. As the British cotton industry in

Lancashire expanded in the nineteenth and early twentieth centuries, this process was observable. More and more enterprises entered the field. The degree of specialization among them became greater and greater; some firms would restrict themselves solely to the spinning of yarn, others to weaving; and in a still later stage an individual firm would restrict its weaving to certain grades and qualities of cloth. As this process continued, each firm was concentrating more and more on a single specialty and was thus able to produce this particular specialty more and more cheaply. The economy occurred, it is true, not so much through the addition of new firms as because the improved organization of the industry permitted reductions in cost to accompany growth in output.

Second, the industry may afford a larger market to machinery manufacturers, raw material producers, and other subsidiary industries, which will lower the cost of purchased materials and services. Finally, the growth of the industry may make possible the development of a more highly skilled and efficient labor force.

Under these conditions an increase in the output of the industry would go along with a decline in price. We must not assume from this that the industry, contrary to the law of supply, has a downward-sloping supply curve for its products. Actually there is no such thing as a downward-sloping supply curve. Barring such exceptional cases as the products brought in by improvident savages, who quit work as soon as they have received a certain amount for their products, how could a lower price possibly induce sellers to offer larger quantities, or a higher price call forth smaller quantities?

In the decreasing-cost industries, a lower price does not cause larger output. The causal relation goes, on the contrary, from quantity produced to price. It is correctly described by saying that if (perhaps through an initial increase in price) output is expanded and costs thereby lowered, the industry can be in equilibrium, after the expansion has occurred, with a lower price than before. But at any given time the industry will conform to the normal law of supply.

Effect of invention. The condition of decreasing cost which we have just examined is to be distinguished from decreases due to technical inventions and mechanical improvements. Technical change is mainly responsible for the fact that the cost of most manufactured products declines fairly steadily over the course of

time. A chart showing cost changes in an industry over a period of time will usually (allowance having been made for changes in the value of money) show a marked downward slope. At any one point of time, however, the supply curve of such an industry will be either horizontal or rising. The supply curve is constructed on the assumption that production methods are fixed; it will shift downward with any improvement in techniques.

This may be illustrated by a graph, as in Figure 20, bearing

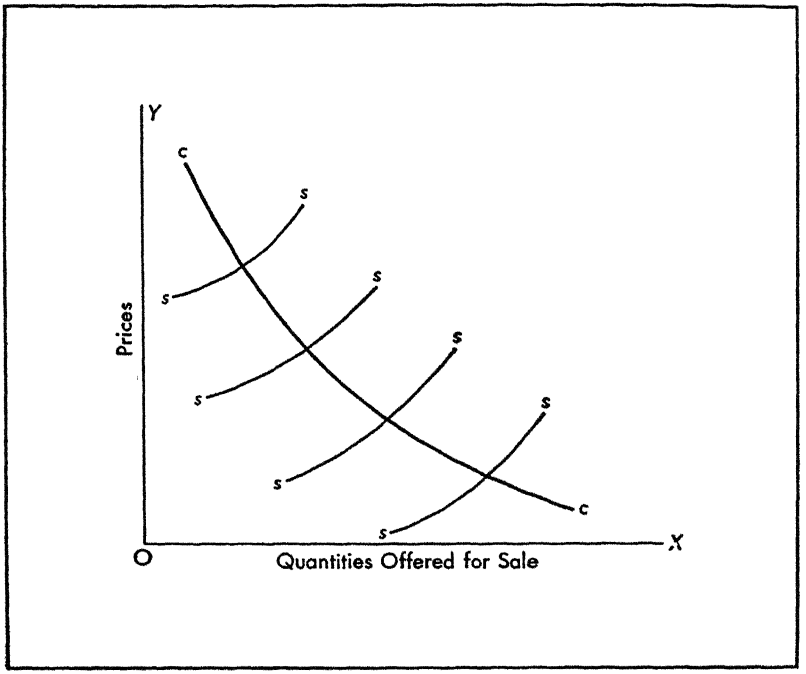


FIG. 20

in mind that the several *ss* curves represent the supply curves of this industry at successive dates and that the curve *cc* portrays a historical sequence, rather than a set of simultaneous possibilities at a given time, as in the conventional supply curve.

EXERCISES

1. Assume the following conditions of cost for the production of automobile heaters in the Acme Products plant:

Output	Fixed cost	Variable cost	Total cost	Average variable cost	Average total cost	Marginal cost
10,000	$150,000	$ 75,000	$225,000	$7.50	$22.50	$ 7.50
20,000	150,000	145,000	295,000	7.25	14.75	7.00
30,000	150,000	216,000	366,000	7.20	12.20	7.10
40,000	150,000	288,500	438,500	7.21	10.96	7.25
50,000	150,000	363,500	513,500	7.27	10.27	7.50
60,000	150,000	442,500	592,500	7.38	9.88	7.90
70,000	150,000	527,500	677,500	7.54	9.68	8.50
80,000	150,000	619,000	769,000	7.74	9.61	9.15
90,000	150,000	719,000	869,000	7.99	9.66	10.00
100,000	150,000	832,000	982,000	8.32	9.82	11.30

(*a*) Plot on a sheet of graph paper the curves of average variable cost, average total cost, and marginal cost.

(*b*) Approximately what is the optimum size, or point of maximum efficiency, for this concern?

(*c*) Assume that the product is sold under conditions of pure competition. If the best market price obtainable for the output were $7.10 per unit, would you expect the concern to produce any heaters? If so, how many? Explain.

(*d*) If the best market price obtainable for the output were $8.50 per unit, would you expect the concern to produce any heaters? If so, how many? What would be the loss per unit, and the total loss, sustained? What would be the loss if the plant closed entirely?

(*e*) If a price of $10 per unit could be obtained for the output, how many heaters would be produced? What would be the profit per unit, and the total profit, under such conditions? Draw the rectangles which correspond to total cost, total receipts, and total profit in this case.

(*f*) Account for the rise in average total costs for outputs above 80,000.

(*g*) Indicate on the graph the firm's supply curve for automobile heaters.

2. (*a*) The fixed costs of the Acme Products plant are estimated above to be $150,000. What in general would have been the effect on the answers to Exercise 1 if the fixed costs had been (1) $100,000? (2) $200,000?

(*b*) If on the other hand, variable costs were estimated to be higher by 50 cents per unit throughout, what would be the effect on the answers in Exercise 1?

CHAPTER XI

Seller's Policy Under Imperfect Competition

We must now go back to the fork in the road at which we had arrived at the beginning of Chapter VII, and setting our feet on the second of the alternative routes, proceed to investigate more fully the nature of imperfect competition and the special forces that determine price under conditions of imperfect competition.

The conditions of pure competition: Résumé. We have already defined pure competition as that type which prevails when the individual seller is unable by his own action appreciably to influence the market price. The conditions that give rise to this type of competition have been summarized as follows: (1) the article dealt in is a standardized product; all units offered by the different competitors are, in the minds of the buyers, exactly alike, so that a lower price is the only reason for any buyer preferring to deal with any particular seller; (2) the number of sellers is very large; (3) the quantity that can be offered for sale by any seller is a very small part of the total quantity offered, so that no change in the amount he offers can have an appreciable effect on the price; (4) all the sellers and buyers are well informed and keen and acting with complete independence of each other.

It will be recalled that under pure competition the individual producer is faced with a horizontal demand curve for his product: he can sell as much as he wishes at the price prevailing in the market. He has no problem of setting a price on his goods or finding buyers for them. His problem is simply to decide how much it will pay him to produce and sell at the price prevailing in the market. Over the price itself he has no control.

The conditions of imperfect competition: Ignorance, inertia, and agreement. Imperfect competition exists whenever one or more of the conditions of pure competition is lacking. Let us start with the last-named of these conditions; *i.e.,* all sellers and buyers are well informed and keen and acting independently. It will be recognized that there are many markets in which there is not that com-

plete knowledge and keenness in buying that furnishes one of the requirements of pure competition. The market may not be well organized, making it difficult for the buyers and sellers to come together and compare each other's wares and prices.

Competition in the retail business is often imperfect to a pronounced degree, because it is so much trouble for buyers to get around and find out exactly what each store is charging for each and every article. Furthermore it is often difficult for buyers to determine how the quality of goods offered by one store compares with the quality in other stores.

Competition in the wholesale business and in the great speculative markets conforms much more nearly to the analytical idea of pure competition, chiefly because the buyers are better informed and more energetic to seize any advantage.

Competition is imperfect whenever sellers and buyers are not acting with complete independence. In actual practice, there may be all degrees of independence. At one extreme, buyers and sellers are acting independently in every respect. In other cases they may be acting independently in some matters but not in others.

The sellers may, for example, agree upon prices, while maintaining complete independence as to advertising and other sales-promotion activities. Or they may compete as to prices, but agree to divide the territory or to limit sales to quotas, etc. As the area of agreement expands, we approach the extreme of complete combination, at which there is agreement on everything.

Complete competition and complete combination are thus the two poles. At the one extreme, we have pure competition—provided the other requirements are also met. At the other extreme, there is a condition of monopoly.

Between these extremes there may be all degrees of partial competition and partial combination. Like many other such practical situations, it may be difficult in some of the middle cases to decide whether competition or combination is the word that applies.

Individuality of products. The requirements of pure competition may be lacking also because the products of competing sellers are not precisely alike. It is typical of the field of industry, as opposed to that of agriculture, that the products of particular producers have considerable individuality. The individual producer is not selling grade A wheat or even a standard quality of cement; on

the contrary, he is selling a particular kind of machinery or a particular brand of beer or a particular make of automobile.

Wherever this is true, individual consumers will have preferences for one make or another. Each producer will have a number of regular and loyal customers who can be depended upon to buy his product even if its price is somewhat higher than others of similar quality. If the price of one make falls a very little below that of others, all buyers will not desert the other makes in a body and concentrate on the one with the slightly lower price.

In the automobile industry, for example, there are many peculiarities in detail among the three close competitors: Ford, Chevrolet, and Plymouth. It is obvious that the demand for Ford cars will depend very directly upon the price that is being charged for Chevrolets and Plymouths. But it is unthinkable that a moderate price cut by Chevrolet and Plymouth would completely deprive the Ford Company of its customers. In this respect the position of the automobile manufacturer offers a sharp contrast to that of the wheat farmer.

Even though the products of different manufacturers may be virtually identical, the sellers frequently succeed in creating the impression of differentiation in the minds of the buyers, especially in the retail markets. Each sugar manufacturer puts up his product in a different package and claims for it distinctive advantages. Canned goods have various labels and various contents. Meats, clothes, gasoline, and very many other everyday products in modern America are offered by each competing producer in a form which he tries to individualize so that it will be known as his own. Brands, forms of packaging, distinctive coloring, slogans, all sorts of advertising are some of the devices used to gain this effect and so to attract a particular following among the buying public.

The basis for buyers' preference may lie also, not in characteristics of the product itself, but in services provided by the seller with the product. The box of corn flakes of a particular brand sold at a grocery which provides credit and delivery services is not the same thing as the same maker's box sold at a cash-and-carry store. Even among cash-and-carry stores, Mrs. Brown may prefer one above others because it is conveniently located, or because the store manager is obliging, or for a dozen other reasons. If this attachment is sufficiently strong, she will buy at the preferred store even though the same good might be had more cheaply elsewhere.

Consciously or otherwise, each individual store develops for better or worse an individuality in its offerings, even though it sells the same brands as another store across the street or around the corner or in the next block. What people pay for in a store is not just merchandise but also the service that goes with it, the general atmosphere of the place, and the convenience of its particular location.

These are some of the things which tend to make competition not pure. A store which raises its prices above those of its competitors will lose some of its customers, and as prices are raised higher and higher, more and more customers will shift their patronage. Conversely, price cuts will attract some, but not all, of the customers from other stores. Different stores will frequently be found charging different prices for the same good, and this is quite consistent with each store securing some patronage.

Small number of sellers. Where there are few sellers of a good, imperfect competition will exist, even though the product is completely standardized. Examples are to be found in such industries as steel, cement, chemicals, and other basic industrial materials. In such industries as these, a single producer may sell up to 10 or 15 per cent of the total sales in a given market. Of course this is not the same thing as 10 or 15 per cent of the total sales in the country.

A structural steel bar meeting certain engineering specifications is the same product whether it comes from Republic Steel, United States Steel, or any other company; and the expert buyers of steel bars are fully aware of this fact. All producers must, therefore, charge the same price, just as they would have to do under pure competition.

In other respects, however, the market for steel products differs markedly from a purely competitive market. Each firm sells to a relatively stable clientele. It cannot, as it could under pure competition, sell an unlimited amount of goods at the prevailing price. Nor can it sell an unlimited amount by reducing its price slightly below the price charged by other producers. Such a reduction would be discovered very quickly by other sellers, who would immediately cut their prices also. Each firm would then be in about the same relative position as before, though all would have larger sales because of the general reduction in price.

If, instead of cutting the price, a firm undertakes to increase its

sales by means of an intensive campaign of advertising and sales promotion, the effect will be similar. Competitors will take notice and redouble their own sales efforts. Total sales may be increased for all, but profits may be reduced because of increased costs.

It is of the very nature of such an industry that no one producer can expand his output very greatly without being willing to sell at lower prices, or to incur much heavier selling costs, or both. The individual producer is too large a part of the particular market in which he sells to leave it unaffected by the quantity of which he disposes.

Large offering of one seller. Failure to meet the requirement that each seller's offering must be a very small part of the total quantity offered has been sufficiently noted as part of the foregoing discussion of number of sellers.

Imperfect competition defined. When competition between sellers fails to meet one or more of the conditions of pure competition, it is customary to say that competition is *imperfect.* We note in all three forms of imperfect competition one common feature; *i.e.,* the individual seller now has some power over the price of his product. Whether because his customers are ignorant or careless or lazy, or because his product has distinctive features marking it off from that of his competitors, or because, as one of a few sellers, his product represents a considerable portion of the total offered for sale, the seller has some discretion in the setting of his own price, and he has some power of retaliation against prices set by his competitors. This is the basis of the definition: *Imperfect competition is competition between sellers under such conditions that the individual seller is able by his own actions to exert some influence upon the market price.*

This definition adapts itself readily to graphical illustration, as in Figure 21. Let us suppose we have an automobile manufacturer who comes directly into competition with only three or four other makes of car. The average price of these other cars is indicated in Figure 21 by the distance *pa.* The number of automobiles he puts on the market is determined, not by price alone, but also by his interpretation of the conditions of demand. He does not have to charge the same price as rival manufacturers. He can charge a higher price such as $p'a'$, or a lower price, such as $p''a''$. At either price buyers can be found, as shown by the demand curve *dd.*

On the other hand, the presence of these other cars on the market

puts a distinct limit on the height to which this particular manufacturer can successfully raise his price. Because of their presence the demand for his particular car is extremely elastic. If he raises the price to *Od* in Figure 21, he will have no sales. Nevertheless the fact remains that up to this level he can charge any price he wants, in accordance with his own judgment of demand, competitive conditions, and costs of production.

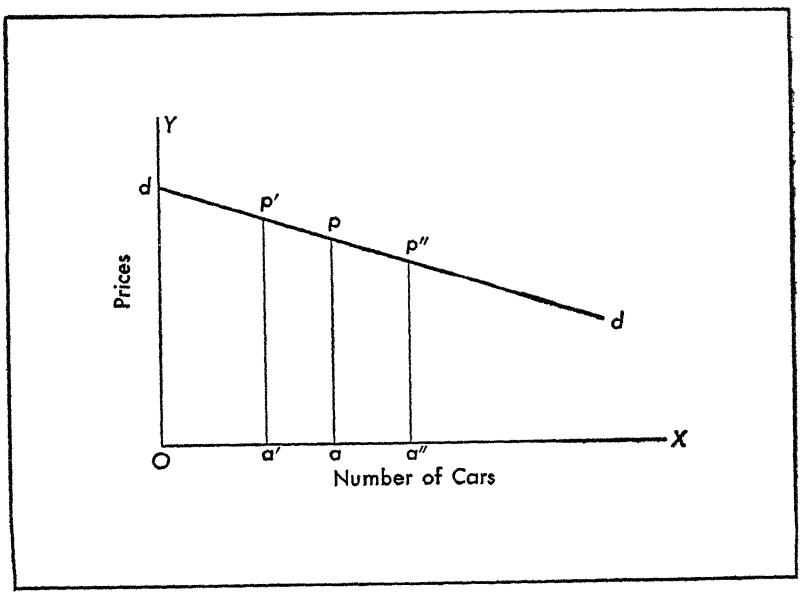


FIG. 21

We need to be on guard against inferring that imperfect competition, in contrast with pure competition, necessarily means a lesser degree of competition, or less active or keen competition, or a certain failure or breakdown of competition. Imperfect competition may be just as complete, active, and vigorous as pure competition. Indeed, there is nowhere more keen and even bitter competition than when on occasion a few powerful rivals are fighting it out. In comparison the pure competition of the wheat farmers seems mild enough. The terms "pure" and "imperfect" signify, not degrees of competition, but different kinds or conditions of competition.

Oligopoly. Where the circumstance that makes competition im-

perfect is the smallness of the number of the sellers, the economists sometimes like to use the term *oligopoly*. This rather forbidding word means, in accordance with its Greek derivation, few sellers. *Oligopoly is a special case of imperfect competition, characterized by fewness of sellers.* The cement market furnishes a good example. The product, cement, is very highly standardized; the ordinary grade conforms so perfectly to rigid government specifications that there is no reason why buyers should have a preference for the cement of any particular manufacturer. Moreover both sellers and buyers in the cement market are remarkably well informed as regards prices and other sales conditions and are keen in the pursuit of their several interests. In these two respects then we have conditions conforming to pure competition. It is only the small number of manufacturers that gives to each seller the power to influence the price of the product.

A classification of market situations. The following classification may possibly be helpful.

I. Pure competition. Standardized product. Many sellers. No agreement
II. Imperfect competition
 A. Many sellers
 1. Standardized product. Agreement
 2. Differentiated product
 a) No agreement
 b) Agreement
 B. Few sellers (oligopoly)
 1. Standardized product
 a) No agreement
 b) Agreement
 2. Differentiated product
 a) No agreement
 b) Agreement
III. Monopoly. One seller or complete combination of sellers

The seller's problem under imperfect competition. When competition is imperfect, the seller has a price problem. He has a certain freedom, but it relates to the seller's side of the market only. The demand for his product is determined by others. Since he usually sells in a free market, he may fix either the price at which he will sell or the quantity. In case he takes the first course,

demand will determine the maximum quantity he can sell. In the other case, demand will determine the maximum price he can charge. In any case, his problem is that of discovering the most profitable price or quantity to set in view of the shape of the demand curve for his individual product and of his various cost schedules, such as we studied in Chapter IX.

Suppose the demand for the product of the firm is as represented by the curve *dd* in Figure 22. *cc* is the representation of the sched-

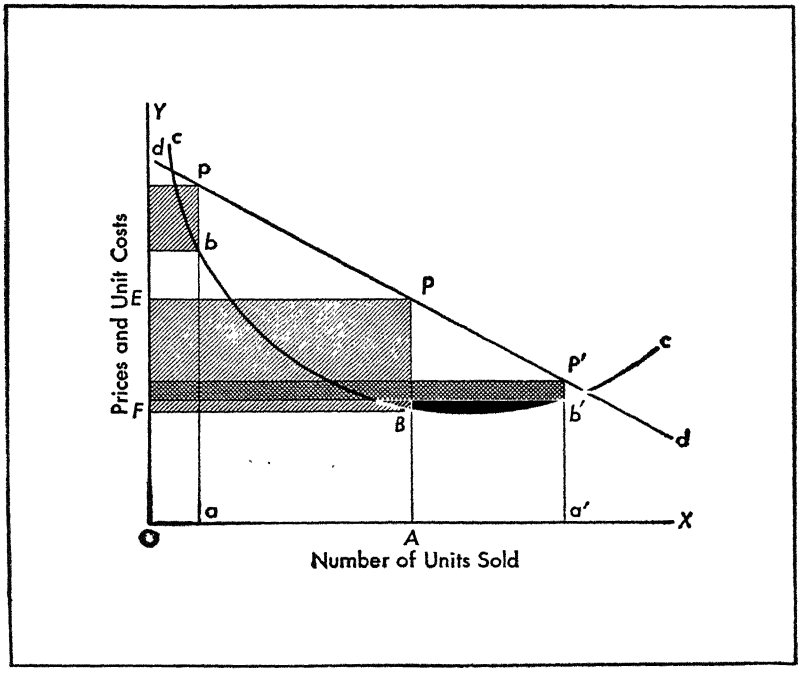


FIG. 22

ule of average total costs. If this producer charges a very high price, such as *pa*, his sales will be only *Oa*. If he hopes to sell the very large amount *Oa'*, he must be prepared to accept the very low price *p'a'*.

In the first case, his total costs are represented by the rectangle with its base on the horizontal axis and its upper right-hand corner at point *b* on the average cost curve, for the area of this rectangle is equal to *Oa* (total quantity) times *ba* (average cost). Under the

same circumstances, his gross receipts are represented by the rectangle with upper right-hand corner at point *p*, this area being equal to *Oa* (total quantity) times *pa* (price). His net profits then are represented by the comparatively small shaded rectangle with corners at *p* and *b*, for this area is equal to the difference between the gross receipts rectangle and the total cost rectangle.

Sizing up the areas of rectangles with upper right-hand corners at *b'* and *p'* respectively, we see that selling the quantity *Oa'* at price *p'a'* would likewise result in comparatively small profit, as represented by the narrow shaded area with right-hand corners at *p'* and *b'*.

But now let us consider an intermediate price such as *PA*. Here the profit rectangle is much greater than in either of the other two cases. If the profit of the firm is to be made a maximum, the problem, translated into graphical terms, is to select a price so that the area of the profit rectangle, *EPBF*, will be as great as possible. In numerical terms, this means finding a price such as $1.00 in the following schedule:

Price	Sales	Gross receipts	Total cost	Net profit
$1.20	6,000	$ 7,200	$ 3,000	$4,200
1.10	8,000	8,800	4,000	4,800
1.00	10,000	10,000	5,000	5,000
0.90	12,000	10,800	6,000	4,800
0.80	14,000	11,200	7,000	4,200
0.70	16,000	11,200	8,000	3,200
0.60	18,000	10,800	9,000	1,800
0.50	20,000	10,000	10,000	0

The question now arises: Is there not some general rule or principle which will describe the attainment of maximum profit? It must be recognized that the producer needs to know the demand for his product. Sometimes his knowledge is very vague—a matter of guesswork. But assuming that he has some idea of the demand for his product, or that he can find out through experiment, how can we describe the most profitable price adjustment that he can make? An answer is found in a very simple relationship between marginal revenue and marginal cost.

Marginal revenue under imperfect competition. We have previously become acquainted with the concepts of marginal cost and marginal revenue, and we have seen how together they determine the policy of the businessman who is selling his product under

conditions of pure competition. The role of marginal cost is the same, whatever the type of competition; but the part played by marginal revenue becomes a bit more complicated when we enter a market of imperfect competition.

This is because the seller in imperfect competition cannot increase his sales without lowering his price. Suppose 90,000 pairs of a certain make of shoes can be sold at a price of $4.50, but to sell 100,000 pairs the price has to be reduced to $4.25. The total revenue or gross receipts from the sale of these respective amounts are as follows:

90,000 pairs @ $4.50	$405,000
100,000 pairs @ 4.25	425,000
Difference in revenue	$20,000

Rate of difference in revenue:
$20,000 ÷ 10,000 = $2.00

Under pure competition, as we know, the seller can sell any amount at the market price. When sales are increased, total revenue is increased by an amount which is the product of the number of additional units sold and the price. In the case of a producer selling in an imperfectly competitive market, however, an increase in sales has to be accompanied by a reduction in price. Thus the change in total revenue is made up of two component parts: it will be increased by the revenue from the sale of the additional units; but it will be decreased as a result of the reduction in price of the number of units which was previously being sold.

Marginal revenue and marginal cost: In general. Whenever a producer discovers that his marginal revenue is greater than his marginal cost, it is clear that he will gain by an increase in production, because such an increase will add more to his revenue than to his cost. As a result of this adjustment, marginal revenue itself will decline and marginal cost may rise, so that a further adjustment will presumably add less to his profits and may eventually even reduce them. But small adjustments which involve a cut in price and consequent increase in sales will always be profitable so long as marginal revenue exceeds marginal cost.

When marginal revenue is equal to marginal cost, so that any further increase in production would increase cost as much as it would increase receipts, additional sales will no longer bring the producer any benefit.

By the same reasoning, if marginal revenue is less than marginal cost, an increase in production would be unprofitable, since it would increase costs more than revenues. Now the only profitable course is to reduce production. In this situation it is to the producer's interest to raise his price (if he sells in a market where price is set by the seller) or directly to limit production.

For the sake of a simple numerical illustration, let us assume that marginal revenue is based upon changes of 2,000 units in sales and that marginal cost will increase gradually with increased product, as in the schedule below.

Price	Sales	Gross receipts	Marginal revenue	Marginal cost
$1.20	6,000	$ 7,200	—	—
1.10	8,000	8,800	$0.80	$0.50
1.00	10,000	10,000	0.60	0.51
0.90	12,000	10,800	0.40	0.52
0.80	14,000	11,200	0.20	0.53
0.70	16,000	11,200	0.00	0.54
0.60	18,000	10,800	−0.20	0.55
0.50	20,000	10,000	−0.40	0.56

Suppose the producer considers selling 6,000 units at a price of $1.20. It pays him to increase the amount to 8,000 units, even though the price has to be reduced to $1.10. His gross receipts with this reduction in price increase $1,600, *i.e.*, at the rate of 80 cents for each additional unit. This more than covers the marginal cost of 50 cents. For similar reasons it pays him to increase production from 8,000 to 10,000 units. In this instance marginal revenue is 60 cents as compared with marginal cost of 51 cents.

It does not pay him, however, to reduce the price to 90 cents and sell 12,000 units. Here marginal revenue is only 40 cents, while marginal cost is 52 cents. If price were reduced to this figure, the final units sold would be adding 12 cents less to revenue than to cost. To be sure, the producer might still be making a profit, but not so much as he could make.

On the other hand, selling 10,000 units at a price of $1.00, he has a marginal revenue somewhat greater than marginal cost. If he were disposed to alter his sales by smaller jumps, he would probably find the ideal output around 11,000 units. With finer calculations of marginal items, he would be able to find an exact equality between marginal revenue and marginal cost at somewhere near this number.

Thus we reach the conclusion that the producer attains his best position when his price policy is such that marginal revenue and marginal cost are equal. This is true, whether the producer is actually making profits or incurring losses.

Maximum profit. It is obviously only when price is above average cost that profits are possible. Let us consider this situation, with the aid of Figure 23. Here *cc* is the average cost curve for a

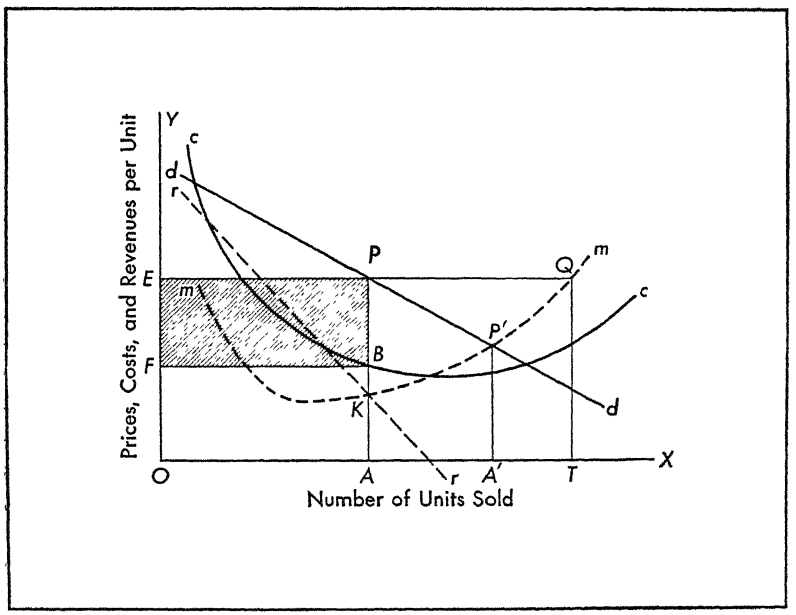


FIG. 23

product sold in an imperfect market; *dd* is the demand curve; *rr* is the representation of a schedule of marginal revenues. The possibility of profit is shown by the fact that the demand curve is above the average cost curve throughout a considerable part of its length.

The marginal revenue schedule is derived from the demand schedule. The curves *dd* and *rr* thus correspond; both relate to the same conditions of demand. The firm's marginal cost curve is *mm*. It intersects the curve *rr* at point *K*[1].

[1] If the marginal revenue curve intersects the *descending* branch of the marginal cost curve instead of its ascending branch as in Figure 23, the intersection still indicates the most profitable volume of operations. The only proviso is that

The perpendicular distance *AK* represents the marginal revenue derived from the sale of the quantity *OA*, and also the marginal cost of this same quantity *OA*. Here marginal revenue and marginal cost are equal.

It pays the firm to expand its operations up to this point. The additional revenue from any small final increase in production covers the additional cost. It does not pay the producer, however, to produce and sell more than *OA*, for, as shown by the two curves, additional cost in this case will not be covered by additional revenue.

With the conditions of demand as represented by the curve *dd*, it is clear that the quantity *OA* can be sold at the price *PA*, determined by extending the perpendicular *AK* up to the demand curve at *P*. *PA* is the price the producer will charge, and the price at which his profits, shown by the shaded rectangle *EFBP*, will be at the maximum.

Minimum loss. In the example just studied, although the firm had to produce the output *OA* in order to maximize its profits, there was still the possibility of earning some profit at a considerable range of other outputs; *i.e.*, at any output for which price exceeded average cost. But not all firms are in this favorable position. In order to round out our analysis of the behavior of producers under imperfect competition, we must consider those firms which are unable at any output to make a profit, because price will always be below average cost.

This condition is illustrated by a demand curve, as *dd* in Figure 24, which throughout its whole length lies below the average cost curve, *cc*. This implies that the firm might lose more money for some outputs than for others, but that it could not under any circumstances cover its costs.

A moment's consideration will make it clear that, although the firm in this situation has no possibility of maximizing profits, its losses may be minimized by equating marginal cost and marginal revenue. If the firm reduces its output below the point at which marginal revenue and marginal cost are equal, then it will reduce its revenues by more than its total costs; consequently its losses will be increased. If it expands output into the region where mar-

the marginal revenue curve must be higher than the marginal cost curve to the left of their intersection. Otherwise the intersection would indicate the least profitable volume of operations, not the most profitable.

ginal cost exceeds marginal revenue, then more will be added to costs than to receipts, and again losses will be increased. We may safely conclude then that, so long as the firm remains in business, it will tend to produce the output at which marginal cost is equal to marginal revenue, so as to minimize its losses.

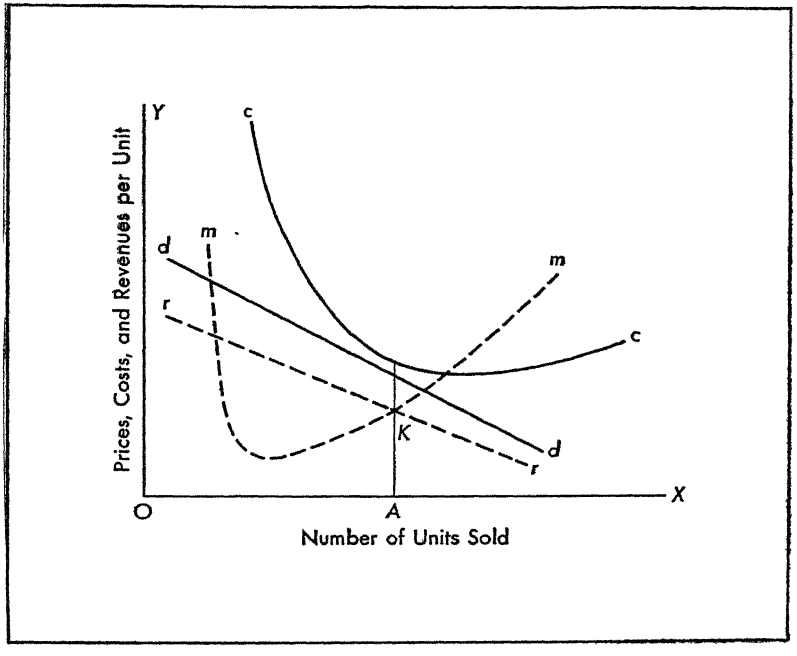


FIG. 24

As to the rest, the situation is similar to that under pure competition. A firm operating under imperfect competition may remain in business for a considerable period of time, operating at a loss, if variable costs are covered; that is to say, if it is earning something, however small, toward meeting its fixed costs. But it cannot thus continue indefinitely.

Conclusions. We may now sum up the results of our investigation of the relationship between cost and the price and output policies of a firm in three propositions, which are analogous to those which emerged from our study of pure competition: (1) So long as a firm continues in operation under imperfect competition, it will tend to produce that output at which marginal cost is just

equal to marginal revenue, thus either maximizing its profits or minimizing its losses, as the case may be; (2) The firm will not continue production in the long run unless it receives a price that is sufficient to cover average total costs; (3) It may continue operations in the short run so long as price exceeds average variable costs.

The seller's liberty under imperfect competition: Pure and imperfect competition compared. Although the seller's problems are more complicated when he sells under conditions of imperfect competition, another characteristic of such a situation is the comparative degree of liberty enjoyed by individual competitors.

This statement may at first appear somewhat paradoxical, because it is the competition between sellers under these very conditions that is usually thought of as being the most vigorous. Farmers never advertise or send salesmen into one another's territory. The economic misfortunes which befall them are usually thought of as attributable in some impersonal way to the market.

But when two or three manufacturers sell their products in a market where there are no other sellers, the competition among them takes a very different form. If producer A increases his sales at the expense of producer B, it is clear to everyone what is the source of producer B's difficulties. Because the number of producers is usually much smaller than under pure competition, the competition among them takes on a more personal form and is likely to be more bitterly resented.

Nevertheless in such a market the individual seller has a freedom which does not exist under pure competition. Unlike the wheat farmer, he may raise his price, perhaps substantially, without losing all of his sales. On the other hand, he cannot expand his sales indefinitely without lowering his price or else spending more money on advertising, salesmen's commissions, etc. As compared with the seller in pure competition, he is in a position where he has greater freedom to raise his price and where the penalties of raising it are less. It is, however, much harder for him to expand his sales, and the cost of so doing is certain to be greater.

Limitations upon the seller's liberty. This liberty of raising price, however, may be so restricted as to be almost negligible, or not worth using. Apparently a businessman may have few rivals in his particular field; actually and indirectly, he may have many. Thus in a particular town there may be only very few stores selling a given kind of product, and competition may appear to be solely

a rivalry among themselves. All of these stores however may be competing, not only with each other, but also with mail-order houses and with larger stores in a near-by city.

True it is that not every shopper has time or inclination to price the article he wants in every shop before he buys it, but some people do go to much trouble in this direction. It is true that every alert businessman tries to individualize his product and his particular type of service. He may thus succeed in establishing more or less permanent business relations with some of his customers. But at the same time a good deal of his trade may be with casual buyers who care no more for him than for any of his rivals. To them one particular brand of merchandise is about the same as several others, and several possible places of purchase may be equally convenient. In spite of the fact that he could raise prices and still do business, the man with established customers often finds it advisable to keep prices down to a freely competitive level in order not to lose the business of alert and independent people who can be easily attracted elsewhere by lower prices.

Liberty and discretion in big business. In the matter of price it may appear that the larger business enterprises have a greater degree of liberty than is enjoyed by small businesses. Nevertheless there is need for discretion, even in the management of an outright monopoly. There is the restraining influence of possible and actual substitutes, a risk which each large concern shares with its rivals, there is the danger of additional competition, the danger of stifling future demand by too high a price in the immediate present, and the possibility of all sorts of popular and political outcries. In addition, the large concern which is not a monopoly is held in check by its competitors. It may still have left a considerable degree of liberty in its price policy, but this liberty has to be used with caution.

It is noteworthy that, in those industries where competitors apparently have the greatest individual freedom, prices are sometimes subject to the most conservative adjustments. A big concern, though bound to no particular price by external conditions, is often forced, by the very complexity of its own internal organization, to adhere to a relatively stable price policy. Price changes involve various notifications to the public, the instruction of numerous employees, and all the red tape characteristic of a big organization. They also complicate the bookkeeping.

It is no wonder then that the management of any big concern, engaged in multitudes of little transactions with the general public, prefers to make as few price changes as possible. Imagine the confusion which would result if Ford or General Motors changed Detroit f.o.b. prices every day, or if retail prices in a big department store were subject to hourly fluctuations. The relative stability of prices in industries where competition is concentrated in few hands is, *ipso facto,* not a demonstration of secret agreements or of arbitrary monopoly power, but rather an inevitable result of the complexity of such enterprises and the need for conservative managerial discretion.

Quality and price in competition. The unsettling results of too many changes in prices have a strong tendency to lead large concerns to base their competitive appeal on quality, salesmanship, and advertising, rather than price change. Instead of selling the same car for less money, automobile manufacturers offer more car for the same money.

Businessmen frequently prefer this sort of competition. It calls for alertness and ceaseless activity and the expenditure of large sums of money in advertising and technical research, but the efforts which each competitor thus puts forth in his own behalf do not provoke the same quick and mutually injurious retaliation as would outright price cutting. For example, if a tire manufacturer has some degree of success in a selling campaign in which he stresses a particular process of treating rubber, he may be sure that his competitors will shortly come forth with something of the same sort, but there the thing will stop until some other new idea crops up in the industry. On the other hand, if a price war is begun, no one can tell where it will stop, or how injurious it may be to all competitors concerned.

It should be remembered that none of this appeal to quality can usurp the fundamental role of price. Raising the quality is, in the final analysis, equivalent to lowering the price. If the quality of a certain make of shoe is so improved that the shoe, though still sold at the former price, will wear twice as long, then one pair will suffice in place of two of the former quality. This is obviously the equivalent of halving the price of a unit of satisfaction.

All competition necessarily involves both price and quality. If the product is standardized and all incidental services are identical, then price is the only means of appeal. If, on the other hand, com-

petitors make their appeal through differentiation in quality, then we simply have one of the conditions making competition imperfect. But the price of the product is no less important. And the buyer of the superior article is getting the equivalent of a lower price for the unit of satisfaction.

The seller needs the facts of demand. Surely no further demonstration is needed of the vital interest of the ordinary businessman in the demand of consumers for the product he makes or sells. Owing to the wide prevalence of imperfect competition, knowledge of the nature of the demand for his goods is a prime qualification of management. The businessman is aware that, other things being equal, the higher his prices, the greater the income and the profits of his business; but his knowledge of the law of demand tells him that other things are not equal, that a high price may defeat its end by reducing sales, while the true road to profit may be large sales stimulated by low prices.

The businessman would like to know the precise form of the demand curve of his product. What will be the response of the buying public to his price policy? Are people already using about as much as they want, or is there possibility of bringing in a crowd of new users by the announcement of lower prices? Is any large part of his present custom likely to drop his product if the price is raised, or may he count upon a condition of inelastic demand?

It is of even more importance to the businessman to know the absolute demand for his products; within any given range of prices are people ready to take large or small quantities? The fate of his business may very well turn upon his knowledge or ignorance of the amount of his goods which buyers would take at various prices. And it should be noted that the entrepreneur wants knowledge of the state of demand, not only as it is at the present moment, but as it will be in the future, both immediate and more remote.

Ascertaining the facts. The task of discovering the actual facts of demand is anything but easy; it is in truth perhaps the most difficult part of the technique of modern business. In spite of the difficulty, however, the importance of the knowledge is such that vigorous efforts are made to obtain facts throwing light on the conditions of demand.

Demand is a resultant of two factors, desire and the purchasing power to make desire effective. The attempts of businessmen to measure and foretell demand revolve about the analysis and

interpretation of business statistics which bear upon these two factors.

To find out the desires of consumers, the businessman avails himself of the statistics of consumption of various products, showing both amounts consumed at particular times and trends. For example, the per capita consumption of beef declined from 67.8 pounds in 1900 to 58.0 pounds in 1927, while the consumption of pork was increasing from 64.7 pounds to 68.5 pounds per capita.

Ability to buy is indicated by regularly published statistics of incomes, showing the sums paid in any year for salaries, wages, pensions, dividends, interest, rents, royalties, etc., and the corresponding indexes showing trends.

A longer range indication of future demand is provided by study of the flow of goods from producers to consumers. The flow from manufacturers, farmers, etc., to the wholesale and retail markets is indicated by statistics of car loadings, exports and imports, and wholesale trade. The flow of goods from these markets to the consumers is shown by such measures as department store sales, life insurance, and advertising. Going still further back toward the sources of income, the businessman studies the records of production, such as the statistics of crop production, receipts of goods at export ports, building construction, etc.

These data are supplemented by indexes of general banking and credit conditions, stock exchange transactions, postal receipts, electric power production, business failures, velocity of bank deposits, new incorporations. All these facts contribute to the picture of general economic conditions and so help businessmen to form their estimates of the future demand for their products.

After a business executive has estimated the future demand for a particular product, he has still to estimate how much of this demand his company will be able to capture. The Goodyear Tire and Rubber Company is interested in knowing whether total tire sales for the coming year will be forty million or fifty million. It is also interested in knowing whether Goodyear will be able to make 20 or 25 or 30 per cent of those sales. A careful analysis of actual sales of Goodyear tires and of rival brands will indicate past trends in "consumer preference," but projection of these trends into the future is very difficult. While a company's percentage of sales is still rising, it may nevertheless be suffering a loss of con-

sumer preference which will be reflected in the sales figures six months hence.

In an effort to detect shifts in consumer preference as nearly as possible, many companies conduct regular preference surveys, either on their own account, or through a market research organization. Many different techniques are used in these surveys. A common method is to telephone a representative sample of consumers in the area being studied and ask them a single question, such as: "What do you think is the best tire made, regardless of price?"; or, "If you were going to buy a tire tomorrow, what brand would you buy?" Not all of those who say they prefer a particular brand will actually buy it when the time arrives. But if the sample is properly selected, the question properly phrased, and the survey repeated at frequent intervals, the company is able to form some opinion of whether its market position is improving or deteriorating.

Sales promotion. The modern businessman is aware that demand may be subject to outside influence, and when such is the case with respect to his own product, he does not hesitate to do his utmost to modify demand in his own interest. To this end the businessman employs a sales force, large or small as the case may be, to whom he pays salaries and commissions; retailers are given special "margins" as inducement to "push" a particular maker's product; exclusive agencies are set up; and finally there is the whole field of advertising. When a survey indicates that consumer preference for a company's brand is declining—*i.e.*, its demand curve is falling—the company need not conclude that production should be curtailed, or even that the product should be altered. It is more likely to conclude that increased selling efforts are required to arrest or reverse the decline of the demand curve.

We should of course not overlook the giving of information to consumers, which is one of the important functions of advertising. Advertising thus frequently makes a real contribution to the process of producing and placing in the consumers' hands the goods that satisfy human wants.

Whether sales promotion efforts such as these are worth attempting or not depends, in the first instance, upon the kind of competition in the particular market. When a product is sold under conditions of pure competition, it will never pay individual producers to employ advertising and other devices to influence con-

sumer taste. It would obviously be ridiculous for a Montana wheat farmer to advertise in a national magazine and urge consumers to "use more wheat and eat more bread." The absurdity arises from the fact that he as an individual is supplying so tiny a fraction of the quantity of wheat consumed in the United States that, even if his advertising increased the sale of wheat a hundred thousand bushels per annum, the effect on his individual sales would be negligible. But when a producer is one of a small number of sellers in the market or is the seller of an individualized product, he may find it well worth his while to advertise.

This is one of the salient features of imperfect competition. We have noted before the impersonal character of most markets characterized by pure competition; this absence of advertising under conditions of pure competition is one aspect of that impersonality. We have noted also that the seller who faces imperfect competition, unlike the seller in a perfectly competitive market, has a price problem. We now see that his situation is further complicated by a sales problem. He will have to decide not only what is the most profitable price to set, in view of the fact that a lower price will expand his sales and a higher price will check them, but also how much money he is going to spend in pushing the sales of his product.

The statement that, in a market characterized by pure competition, advertising does not pay refers of course to the individual seller. It does not follow that, even in case of pure competition, it may not be profitable for many sellers to engage in a joint advertising campaign. Thus the producers of wheat might conceivably, through a national organization, seek by advertising to increase the demand for wheat, and so its price, to the benefit of each producer. There is a great deal of national advertising conducted by associations of producers which it would not pay any one producer to undertake. In some of these cases, the product is standardized enough and the number of sellers sufficiently great that there is approach at least to a condition of pure competition.

In the second place, it is clear that selling expenditures cannot be increased indefinitely. The seller under imperfect competition is thus faced with an additional problem from which the producer under pure competition is free. He must determine what level of selling expenditures will enable the firm to earn maximum profits. Any cost-and-demand diagram, such as Figure 23 (page 197), assumes a certain level of selling costs.

An increase in selling costs will raise both the average cost curve *cc* and presumably the demand curve *dd*. After a time, however, the effect on the demand curve will begin to diminish, and a point will eventually be reached at which additional selling efforts will have no effect at all; *i.e., dd* will have reached its maximum position. The company should clearly stop considerably short of this point. But precisely what level of selling expenditures will maximize the company's profits?

The rule which guides the seller is very similar to the rule of equalizing marginal cost and marginal revenue, which we discovered in analyzing the proper level of output from a plant. Expansion of selling efforts will be profitable so long as the increase in net revenue—*i.e.,* the increase in sales receipts less any increase in production costs—exceeds the increase in selling costs. When these two quantities become equal, the optimum level of selling costs has been reached.

To put the matter in different words, the firm is confronted, not with the single cost-and-demand situation shown in Figure 23, but with a large number of possible Figure 23's—each corresponding to a different level of selling expenditures. From this array of diagrams, the firm needs to select the one for which the profit rectangle *EFBP* is largest. It is of course assumed that, for each level of selling costs, that rate of output is selected which equalizes marginal cost and marginal revenue and which therefore maximizes *EFBP*. The objective of the company is to secure the largest of these possible maximums.

It is much easier to state this principle than to apply it to an actual business situation. It is usually impossible to predict at all accurately how demand will respond to an increase in selling efforts. Even after the increase has occurred, it may be very difficult to distinguish the effect of the increase on sales from the effects of other changes in market conditions. Moreover, an increase in selling efforts by one firm may stimulate counter measures by rival firms, which further complicates the situation.

Advertising. Advertising has become an enterprise of stupendous proportions. The following is a partial estimate of the amount of money expended on three of the principal kinds of advertising in the United States in the first eleven months of the year 1939.[1]

[1] U. S. Bureau of Foreign and Domestic Commerce, *Survey of Current Business,* Washington, January, 1940, p. 23.

Newspapers (52 cities only)	$1,125,446,000
Magazines	138,698,000
Radio	74,988,000
Total	$1,339,132,000

This is obviously far from a complete account. It does not include such important types as direct advertising, streetcar advertising, billboard advertising, business papers, etc.; only 52 cities are included in the newspaper advertising; and finally the record does not cover the complete year.

Although few businesses are today able to function without resort to advertising, the volume of advertising naturally varies greatly among different lines of industry. "The advertising of automobiles; toilet articles and medicinal preparations; food, groceries, and beverages; and cigars, cigarettes, and tobacco constituted 60.5 per cent of the total national advertising in newspapers in these cities (49 cities of the United States) in 1927." [1]

EXERCISES

1. Assume the following conditions of cost for the production of automobile heaters in the Acme Products plant:

Output	Fixed cost	Variable cost	Total cost	Average variable cost	Average total cost	Marginal cost
10,000	$150,000	$ 75,000	$225,000	$7.50	$22.50	$ 7.50
20,000	150,000	145,000	295,000	7.25	14.75	7.00
30,000	150,000	216,000	366,000	7.20	12.20	7.10
40,000	150,000	288,500	438,500	7.21	10.96	7.25
50,000	150,000	363,500	513,500	7.27	10.27	7.50
60,000	150,000	442,500	592,500	7.38	9.88	7.90
70,000	150,000	527,500	677,500	7.54	9.68	8.50
80,000	150,000	619,000	769,000	7.74	9.61	9.15
90,000	150,000	719,000	869,000	7.99	9.66	10.00
100,000	150,000	832,000	982,000	8.32	9.82	11.30

The student will note that these are the same conditions as were assumed as the basis of Exercise 1 at the close of Chapter X.

(*a*) Plot on a sheet of graph paper the curves of average variable cost, average total cost, and marginal cost.

(*b*) Approximately what is the optimum size, or point of maximum efficiency, for this concern?

2. Assume that the product of this concern is not homogeneous with that

[1] Copeland, M. T., in Committee on Recent Economic Changes of the President's Conference on Unemployment: *Recent Economic Changes in the United States*. New York: McGraw-Hill Book Company, 1929, p. 415.

of its competitors, but is distinguished through branding or special features of construction. The table below represents the demand schedule for Acme heaters, sold under conditions of imperfect competition:

Prices	Quantities
$15.00	10,000
14.00	20,000
13.00	30,000
12.00	40,000
11.00	50,000
10.00	60,000
9.00	70,000
8.00	80,000
7.00	90,000
6.00	100,000

(*a*) For each price and amount of sales in the demand schedule, compute the total receipts that would be obtained from selling that amount at the corresponding price.

(*b*) By reference to the figures of total cost in Exercise 1, compute the net profit or loss at each possible price and output.

(*c*) At what price and output are the profits of the concern maximized?

3. (*a*) From the demand schedule in Exercise 2, compute the figures for marginal revenue.

(*b*) Plot on a sheet of graph paper the demand curve, and the marginal revenue curve.

(*c*) From the cost schedule in Exercise 1, plot again on the graph prepared for (*b*) the curves of average total cost and marginal cost (four curves on the same graph).

(*d*) What is the best output for the concern to produce? Reading from the chart, at what price can this output be sold? What is the average total cost of producing this output?

(*e*) What is the profit per unit and the total profit under these conditions? Draw the rectangle which corresponds to total profit.

(*f*) Describe the degree of elasticity of demand for the Acme heater.

4. Assume now a different demand schedule for Acme heaters, as indicated below:

Prices	Quantities
$10.00	40,000
9.50	50,000
9.00	60,000
8.50	70,000
8.00	80,000
7.50	90,000
7.00	100,000

(*a*) For each price and amount of sales in the demand schedule, compute the total receipts, and the marginal revenue, that would be obtained.

(*b*) Plot on a sheet of graph paper the demand curve, and the marginal revenue curve.

(*c*) From the cost schedule in Exercise 1, plot again on the graph prepared for (*b*) the curves of average total cost and marginal cost (four curves on the same graph).

(*d*) What is the best output for the concern to produce? Reading from the chart, at what price can this output be sold? What is the average total cost of producing this output?

(*e*) What is the loss per unit, and the total loss, under these conditions? Draw the rectangle which corresponds to total loss. What would be the loss if the plant closed entirely?

(*f*) Discuss the elasticity of demand for Acme heaters as disclosed by the demand schedule.

5. The manufacturer of Acme heaters might decide to "increase his sales" of heaters by offering a special inducement to the wholesalers and jobbers in the form of "quantity discounts." As an alternative plan, he might undertake an advertising campaign at, say, a cost of $100,000. Upon which section of his costs, fixed or variable, would these expenditures impinge? Suppose that he anticipates that either of these plans would shift his demand curve to the right by about 10 per cent. How would his most profitable output differ from that indicated in Exercise 4?

CHAPTER XII

Price Under Imperfect Competition. Monopoly

Pure and imperfect competition. We are now ready to study the way prices are determined in an imperfectly competitive market. In Chapter XI imperfect competition was defined and compared with pure competition. We learned that, in contrast with pure competition, imperfect competition does not permit the seller to increase his sales without either a reduction of price or an intensive and costly sales promotion program. The demand curve for his product is not horizontal, as in the case of one who sells under pure competition, but is downward sloping. This situation of a sloping demand curve, more or less shiftable by sales activities, is the characteristic feature of imperfect competition.

Principal bases of imperfect competition. Imperfect competition usually arises from differentiation of the product, fewness of sellers, or a combination of the two. This last condition is quite common. For example, automobiles, tires, gasoline, refrigerators, soap, cigarettes, and many other consumer goods are produced by comparatively few manufacturers, each of whom attempts to stimulate buyers' preference for his own brand.

The consequences of imperfect competition are most readily analyzed, however, by considering markets in which only one source of imperfect competition is present; *i.e.,* either few sellers or a differentiated product, but not both. It will be convenient to consider first the case of a product or service of which many sellers offer differentiated versions. The operation of such a market is similar in many ways to that of a purely competitive market; it will therefore provide a convenient stepping-stone to the more complex relationships among a limited number of sellers.

Many sellers and differentiated product: What do we mean by "many"? The term "many sellers" needs more explicit definition than we have yet given it. How many sellers? Without knowing the circumstances of the market in question, it is not possible to say whether ten sellers constitute "many" or whether the number

should be fifty or two hundred. The test is not numbers as such, but a certain type of behavior on the part of the sellers.

We say that there are many sellers when the number of sellers is so large and the amount that could be offered for sale by each individual seller is so small that a price change by one seller will not appreciably affect the demand curve for the output of any other seller. This being so, a price change by one seller will not lead other sellers in the market to change their prices.

Conversely there are "few" sellers in a market when a price change by any one of them will affect other firms' demand curves sufficiently so that they may change their prices as well, In this case a price change by any firm may set off a general price change throughout the market.

Effect of buyers' preferences. Where there are many sellers, each seller can go his own way on prices without considering the reaction of other sellers to his decisions. He is free to set his price and output at the point which will maximize his own profit. Just as under pure competition, maximum profit is attained at that rate of output which equalizes marginal revenue and marginal cost. The only difference is that, because of the existence of buyers' preferences, the firm's demand curve and marginal revenue curve slope downward from left to right.

The elasticity of the demand for the firm's product indicates the strength of buyer preference for its product. If a baker's customers were strongly convinced of the superiority of his bread and very reluctant to buy elsewhere, even at much lower prices, the demand for his bread would be quite inelastic. If their preference for his product were slight and easily offset by lower prices elsewhere, his demand would be very elastic. Going to the extreme of elasticity, if customers were indifferent to everything save price, the demand curve would be horizontal, and the seller would be operating under pure competition.

Aggregate demand for the commodity being given, the position of the firm's demand curve is determined by the number of other sellers in the market and the prices which they charge. A firm's demand curve will shift downward if new sellers enter the market or if established sellers cut their prices; it will move upward from the opposite causes.

Short-run problem. The short-run equilibrium of the firm may be illustrated by referring back to Figure 23 on page 197, which is

here reproduced as Figure 25. The most profitable output for this firm is clearly *OA*, and the price at which this output can be sold is *AP*. The firm is able to earn net profits equal to the area *EFBP*. It might, of course, be making net losses if the demand curve lay entirely below the average cost curve.

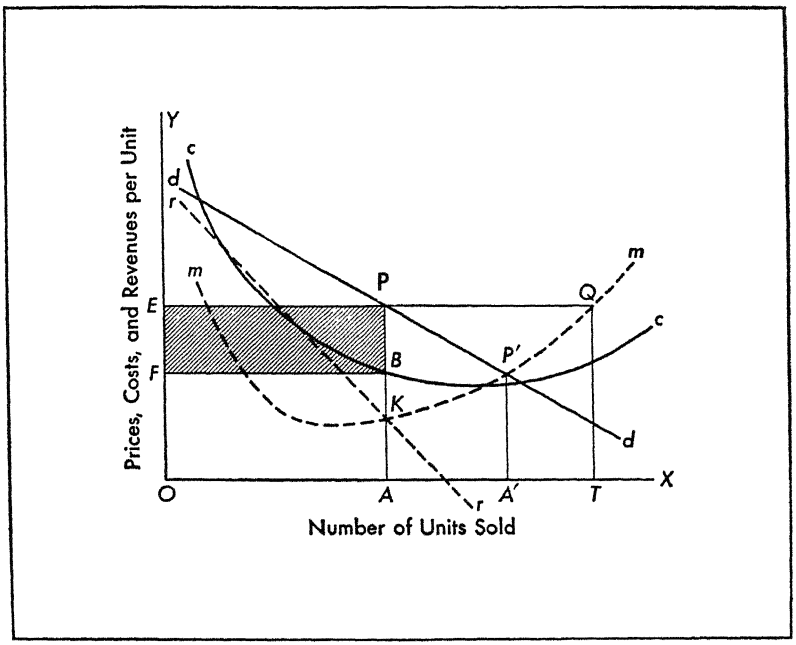


FIG. 25

Long-run problem. Even if the firm is able to make substantial profits in the short run, these profits will not necessarily continue. Unless the industry of which it is a part is organized to prevent the entrance of new competitors, or unless the profits that the firm is earning are expected to be temporary, new competitors will be attracted to the same line of business by the high level of profits. As new competitors enter the field, moreover, some of them may offer closer substitutes for the product of our firm than have hitherto been available.

It is probable, therefore, that the demand curve *dd* may shift downward and will at the same time become more elastic. This process may continue until the demand curve has fallen to the

position shown in Figure 26. By producing the output *OA* and selling it for *AP*, the firm can just cover its average costs. Any other output would involve a net loss.

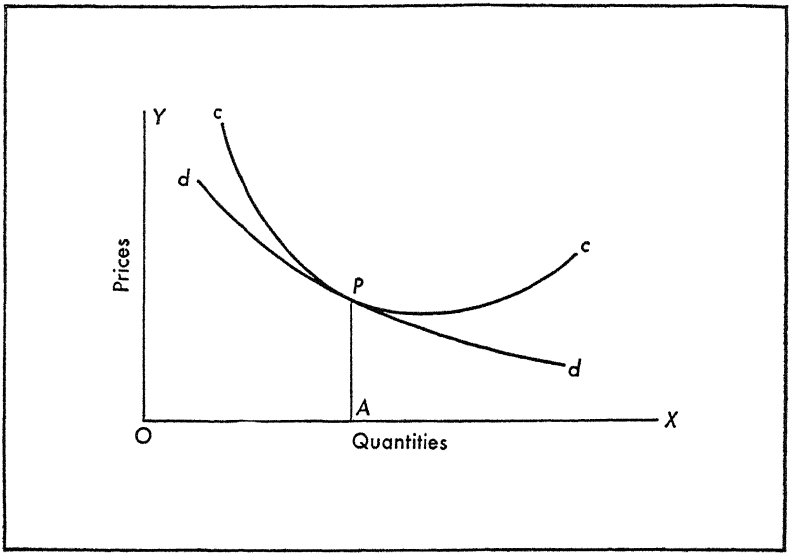


FIG. 26

By the time the firm we are studying has reached this position, the profits earned by firms producing substitutes will also have been forced down to the same level. There will therefore be no incentive for new firms to put still other substitutes on the market. In this particular sector an equilibrium will have been reached, exactly similar to the equilibrium achieved in an industry under pure competition when the entrance of new competitors has driven the market price down to the point where it just covers the average costs of the various firms in the industry.

The conclusion just reached may be stated more precisely as follows: If there are many sellers in a market, if all sellers have identical cost and demand curves, and if new firms are free to enter at will, the market cannot be in full equilibrium until each firm is in the situation shown when it is just able to cover its cost (as in Figure 26). The price *AP* is the long-run equilibrium price in the market, and it will be uniform for all sellers.

In any actual industry, some firms have lower cost curves than

others, some have particularly favorable demand curves, and some are better protected than others against the encroachment of substitute products. Under these conditions, a full equilibrium, in which all sellers are just covering their costs and in which firms have ceased to enter or leave the industry, will not be found.

The nearest thing to it which might be found in actuality is a situation in which firms of average efficiency are just covering costs, with some firms in the industry making net profits and others net losses, and in which the departure of unprofitable firms from the industry is counterbalanced by the entrance of new firms hoping to secure some of the profits of the successful enterprises.

Equilibrium in practice is likely to involve considerable turnover, rather than complete absence of motion. In industries such as the retail grocery trade or the manufacture of clothing, this turnover is very large. In years of normal business about 20 per cent of the manufacturers of women's dresses who were in operation at the beginning of the year are out of operation at the end of the year. The rate of failure among retail grocers is almost as high, and it is substantial, though lower, in other branches of retail trade.

This turnover is not due primarily to the conditions of imperfect competition, since it exists also in purely competitive industries. It results mainly from the fact that entrants are poorly informed about cost and demand conditions and also tend to overestimate their own managerial ability.

This discussion has shown that imperfect competition based solely on product differentiation does not provide any assurance of abnormally large profits. On the contrary, where there are many producers and where new producers are able to enter the industry and provide close substitutes for the products of existing firms, profits cannot, in the long run, be any larger than they would have been under pure competition.

Imperfect competition with few sellers. We turn now to examine price determination in a market where there are few sellers. For the sake of simplicity, let us assume also that the product of these sellers is completely standardized. To introduce product differentiation would complicate the discussion without changing it materially.

The meaning of "few." The term "few," it will be recalled, is not defined on a numerical basis. The criterion is rather a type of competitive relation among the sellers. We say that there are few sellers if each firm's sales are noticeably affected by changes in the price

or sales efforts of other firms, if each firm therefore has reason to expect that any change in its own price or sales policies will provoke reactions from other firms, and if each firm tries to estimate and take account of such reactions before making price decisions.

We may put the matter another way: Where there are many sellers, competition is atomistic and anonymous, and each firm's price and sales policies are an adaptation to impersonal market forces. Where there are few sellers, competition is conscious and deliberate, and each firm's market strategy is an adaptation to the present and prospective strategies of other firms.

Why are sellers so few? It is clear that fewness of sellers is a common situation in actual commodity markets, and the reasons are not difficult to find. In many cases the optimum scale of plant is so large, relative to the aggregate demand for the product, that there is room for only a few plants in the industry.

This may mean either that the plants are large in an absolute sense, or that the market is so restricted that only a few firms can afford to specialize on the product. There are few basic steel producers because of the very large plant required; there are few producers of wire nails or door hinges because of the smallness of the market. Again while the total number of chemical firms is very large, there are usually few producers of a particular chemical product.

Geographical dispersion of buyers and sellers is an important factor restricting the number of sellers in a particular market. If there are only three gravel pits near a town, and if the cost of trucking gravel from elsewhere is prohibitive, then it is of no importance that there are several thousand sellers of sand and gravel in the United States as a whole.

These factors are probably more important than the deliberate restrictions on the number of sellers which are often stressed in discussions of "the monopoly problem." Combination of competing firms for purposes of price control or financial manipulation, exclusion of new firms by patent privileges or monopolizing of raw materials, and similar restrictive devices have doubtless played a considerable part in shaping the present organization of industry in the United States. But the most ideal institutional arrangements for maintaining a free market could not remove those reasons for fewness of sellers which rest on the facts of technology and geography.

Short-run problem: The demand curve. Fewness of sellers complicates the problem of constructing a demand curve for the product of the individual firm. Where there are many sellers, the demand curve can be drawn on the customary assumption that the prices of all other sellers are given. The demand curve then shows the effect on the firm's sales of a change in its price, no change having occurred in the prices of other sellers.

Where sellers are few, such a demand curve, while it might still be drawn, would have little significance. For, by the very definition of "fewness," a price change by one seller will provoke reactions by others, and we cannot assume that other prices will remain the same. A price cut by one seller will most commonly cause all other sellers to cut prices by the same amount. It is possible, on the other hand, that other sellers may cut by a greater or smaller amount, or that some sellers may cut prices and others not.

In any event, the demand curve for the product of an individual firm must be so constructed as to take account of the reactions of all other firms in the market; *i.e.*, it must show the amount which the firm will be able to sell at a particular price after all other firms have adjusted their prices and sales policies to this price. In the commonest case, where a price change by one seller results in equal price changes by all others, each firm's demand curve will have the same elasticity as the aggregate demand curve for the product. It will naturally lie farther to the left, however, its distance from the vertical axis depending on the firm's share of the total market at a particular time.

Limits of price and output. Suppose that *dd* in Figure 27 (page 218) is a demand curve, constructed on this basis, for a firm in a market of few sellers. The curves *rr, cc,* and *mm* have the same significance as in previous diagrams. The curve *vv* has been added to indicate average variable costs per unit of output. These may be taken for present purposes as equivalent to out-of-pocket costs. In practice, out-of-pocket costs would be slightly greater than this, since some types of overhead cost must be paid currently. Suppose further that Figure 27 represents the situation of each of the other firms in the industry.

In the short run, the price charged by each firm will lie somewhere between OE and OE_2, and its output will lie between OA and OA_2. No firm will willingly set its price above OE, since this price and output equalize marginal cost and marginal revenue,

and therefore yield maximum profit. On the other hand, the price cannot fall below OE_2 for any length of time; since this price just equals marginal cost, there can be no reason for further price reductions.

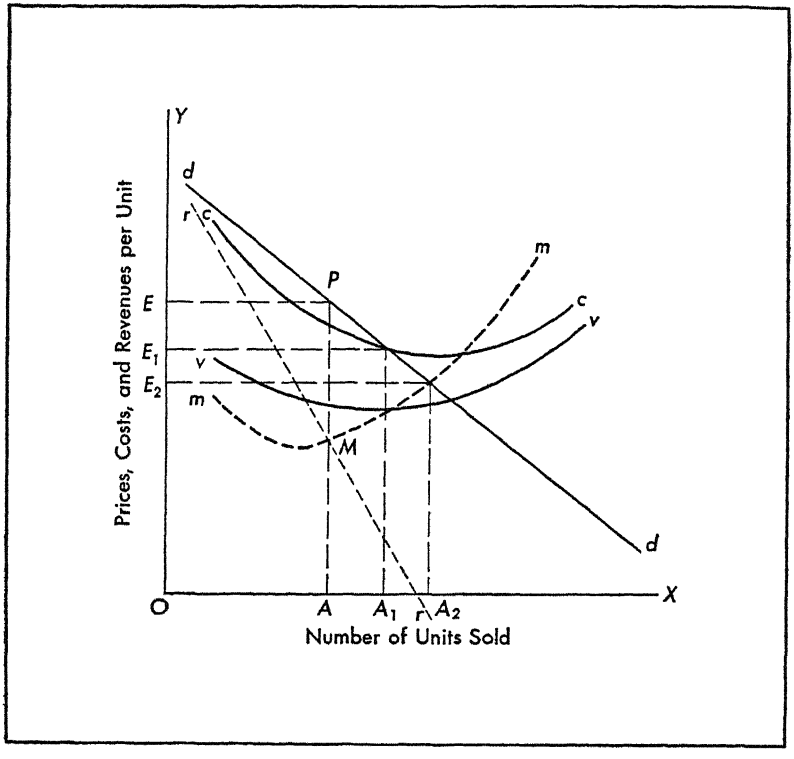


FIG. 27

The seller's interest against price cutting. It is clearly to the interest of each firm in the industry that the price be set as close to OE as possible. If prices are cut below this point, profits will be reduced and may eventually be turned into losses. Once price cutting sets in, moreover, one can never be sure how far it will go. Even if the price happens to be set initially above OE, therefore, it may be considered safer to leave it somewhat too high than to start a process of price reduction whose end cannot be foreseen.

The feeling that price reductions are dangerous is frequently held by businessmen. Terms such as "uncoöperative" and "un-

ethical" indicate the distrust with which the price cutter is regarded. This does not mean that any reduction of prices is opposed under all circumstances. The thing which is condemned is an independent decision by one seller to make a price change unsanctioned by, and unprofitable to, the group as a whole. It is the nonconformist, the man who insists on going his own way regardless of the effect on the industry, who incurs the displeasure of his business colleagues.

Safeguards against price cutting: "Price leadership." The effort to protect group interests against the effects of unbridled individualism has led to the development of such devices as *price leadership,* under which all sellers in a market look to one of their number to initiate price changes and customarily change their prices at the same time and by the same amount. The *price leader* may be a true leader in the sense of making an independent appraisal of demand conditions and an independent decision on the proper level of prices. More frequently, however, the leader acts only after canvassing the judgment of other members of the industry, and his action represents a consensus of opinion, rather than an independent decision imposed on other firms.

Agreements by sellers. The price may also be held at the desired level—*OE,* in this instance—through a definite agreement by all sellers in the market to observe a uniform schedule of prices and to consult with each other before changing this schedule. These agreements vary greatly in formality and complexity; they may be administered through a trade association, a firm of management engineers or accountants, an *ad hoc* committee, or a mere informal consultation.

The formation of price agreements is limited in the United States by the Sherman Act and other antitrust legislation, which, broadly speaking, render such agreements unlawful if they affect interstate commerce. Price agreements are rather common in local trade and service industries, where such legal restraints are generally absent.

Economic obstacles to price control by sellers. More effective than the legal prohibitions, however, are certain economic obstacles to the regulation of prices by groups of sellers. A brief consideration of some of these obstacles will reveal why it is often difficult to establish a price agreement and to maintain it intact over a period of changing economic conditions.

Unused capacity. A very important factor is the unused plant capacity which, except in periods of peak prosperity, is to be found rather generally throughout industry. If a plant is only partially used, it will pay to produce more from it if the extra product can be sold at any price greater than marginal cost.

Consider, for example, the situation of the producer in Figure 27 (page 218). He has succeeded in establishing the price *OE* and is producing an output *OA*. He can increase his profits if, without disturbing the price *OE* on his regular sales, he can somehow dispose of additional output at anything above its marginal cost, as shown by *AM*. He cannot announce publicly that he is willing to do this, for his competitors would then cut prices also and he would slide down the demand curve *dd* to a less profitable position. But he may be able to give a special discount to one or more large buyers, or sell under a different brand name to a chain store at a lower price, or make some other sort of concealed price concession.

The danger of this policy is that even the most carefully concealed price reductions will in time become known to his competitors and will lead to retaliatory price cuts which may undermine the whole price structure of the industry. The temptation of additional orders, however, is frequently sufficient to offset this danger in the seller's mind. This is particularly true in the manufacture of consumer goods, where chain and department store organizations are willing to place very large orders in return for price concessions and can frequently play off one manufacturer against another. These large distributors have played an important role in preventing price coöperation among manufacturers and in holding prices to a lower level than they might otherwise have reached.

Marginal producers. Another factor working in the same direction is the presence in many industries of marginal producers, whose average total cost curves lie entirely above their demand curves and who are therefore unable to avoid losses. They can still reduce their losses, however, if they can secure additional business at anything above its marginal cost to them. Some of these firms, in the desperate hope of avoiding, or at least postponing, bankruptcy, will seek additional business at low prices, regardless of the adverse effect on the price structure of the industry as a whole. It is these frantic efforts at self-preservation, particularly common in depression periods, which are most commonly indicated

by such terms as "chiseling," "undercutting," "cutthroat competition," and so on.

Conflicting interests of sellers. Another source of difficulty in securing price coöperation among sellers is the existence of differences of economic interest among the rival firms, based on differences in their cost and demand situation.

Such differences exist, for example, between high-cost and low-cost firms in an industry. The low-cost firms may, in order to preserve harmony, accept a level of prices high enough to keep the high-cost firms in operation. On the other hand, they may consider it to their interest to force prices down and take business away from the high-cost firms. They may prefer to "cut and expand" rather than to "live and let live." If even one important firm in an industry takes this position, it can force prices down to the level it wishes.

Again, firms may differ in their estimates of the elasticity of the aggregate demand curve for the industry. A firm which believes that this demand curve is very elastic may cut prices over the opposition of other firms because it believes that it is acting in the true interests of the industry as a whole.

Apart from these actual differences of interest, some firms appear to value independence of action for its own sake. This sort of policy, usually resulting from the dominance of a single strong personality—a Henry Ford or a Harvey Firestone—is probably declining in large-scale industry, but it may still be quite important among smaller one-man enterprises.

Conclusion. We return, then, to the conclusion set forth at the beginning of our discussion of the limits of price and output: The firm's price will ordinarily lie, in the short run, somewhere between OE and OE_2 (see Figure 27). The firm may sometimes consent, for strategic reasons, to a price higher than OE, and irrational price cutting may for a short time carry prices below OE_2. In general, however, the conclusion is valid. Precisely where the price will settle within this range cannot be stated in general terms. The result depends on the circumstances previously discussed and on other factors beyond the scope of this chapter.

If a high degree of collaboration is achieved within the industry, prices will be stabilized in the vicinity of OE. If, for one reason or another, consensus cannot be obtained, prices will settle at a lower level. Even in this case they will probably be held somewhat

above the extreme low point, OE_2, by temporary agreements and individual self-restraint. Some industries appear to alternate between periods of price warfare, during which prices fall to the neighborhood of OE_2, and temporary truces, during which prices are raised somewhat above this level.

Long-run problem. We must now consider those factors affecting price which become fully operative only over a period of several years. While price may fall temporarily below the average total cost of the product, it cannot remain there indefinitely. A price lower than OE_1, since it fails to cover total costs, will lead firms to withdraw from the industry. As this happens, the demand curves of the remaining firms will rise, until they are able to secure a price which will cover costs. In the long run, therefore, the price must lie somewhere between OE and OE_1; the range from OE_1 to OE_2 is of only short-run importance.

It should be pointed out, however, that it takes considerable time —possibly as long as ten or fifteen years—for low prices alone to force enough plants out of an industry to restore profitable operation for the remainder. A plant can and will struggle on for a surprising length of time, even when it is doing little more than cover out-of-pocket costs. This is one reason for the bitterness of businessmen's complaints against cutthroat competition. Given sufficient time, competitive forces are capable of adjusting the plant capacity of an industry to the aggregate demand for its products. But the process of adjustment is slow and painful.

Any price higher than OE_1 will offer better than the normal prospects of profits to the firm. Present high profits, however, will not necessarily continue. Just as in the case where there were many sellers, new firms may enter the industry until the demand curves of existing firms fall to a position of tangency to their total cost curves. In this event the point of tangency indicates the only price and output at which producers can cover costs.

It is probable, however, that the barriers to the entrance of new firms are stronger where there are few sellers than where there are many. In a market with few sellers, the aggregate demand, though large enough to offer them a good profit, may not be large enough to accommodate another firm of optimum size. The new firm might shift the demand curve of existing firms so far to the left that all would suffer losses. Moreover in a market with few sellers, any newcomer is bound to be highly conspicuous, and it is natural

that existing firms should try to dissuade him in one way or another —by securing control of the materials which he needs, influencing his sources of credit, pointing out the danger of patent suits if he persists in his plans, or simply buying him out.

These tactics make it difficult for a completely new firm to gain entrance to an industry already dominated by a few well-established producers. It is somewhat easier for a firm already well established in one industry to invade a neighboring industry—for Henry Kaiser to produce automobiles, or for General Motors to produce refrigerators.

Suppose that the firm in Figure 27 is able, in collaboration with its competitors, to maintain a price in the vicinity of *OE*. Suppose further that new producers are effectively restrained from entering the industry. This still does not mean that the firm's profits are safe from attack.

Over a long period demand may shift away from this industry entirely to new substitute products—from carriage wheels to rubber tires, from Pullman cars to passenger planes. Costs may rise because of growing scarcity of some material on which the industry depends. Costs may rise also because of rivalry in advertising and other selling expenditures.

When firms are prevented by agreement or custom from winning away each other's business through price cutting, they can increase their sales only by intensifying their selling efforts. If only one firm in an industry did this, it might hope to raise its demand curve sufficiently to more than offset the increase in selling costs. If all firms increase their selling efforts simultaneously, however, their demand curves will end up in about the same relative position as before. The main result will be higher costs and lower profits all round.

If sellers were perfectly rational, they would probably try to prevent this erosion of profits by setting an upper limit to selling expenditures, as well as a lower limit to prices. Agreements of this sort however are rarely, if ever, found in practice.

Effect of fewness of buyers. We have dealt thus far only with market imperfections resulting from product differentiation and fewness of sellers. Imperfect competition may also arise because there are few buyers in the market. The situation of few buyers and many sellers is common in the markets for agricultural prod-

ucts—notably, livestock, fresh fruits and vegetables for canning, fluid milk and cream, and tobacco.

While the operation of such markets has not been thoroughly analyzed, it is clear that if the buyers acted in concert they should be able to depress the price below that which would prevail in a purely competitive market. Farmers have tended to adjust to this situation by forming coöperative marketing associations. In consequence, while some farm crops are still sold in purely competitive markets, an increasing proportion of agricultural produce is sold at prices determined by bargaining between organized groups of buyers and of sellers.

Large chain and department store buyers hold a strategic position in the markets for manufactured consumer goods. Particularly where manufacturers are numerous and unorganized, these large buyers are able to obtain prices well below those granted to smaller buyers and often only slightly above the manufacturers' marginal costs.

Conclusions. Our study of price determination in imperfectly competitive markets may seem inconclusive and unsatisfactory compared with the clean-cut analysis under pure competition which preceded it. In a market characterized by pure competition many buyers and many sellers trade in a single standardized commodity. Therefore the study of price determination under conditions of pure competition is a study of the way in which the price is determined by market forces. Given a certain demand and supply situation, it is possible to say precisely what price will be established.

Under imperfect competition, the number of variable factors which must be taken into account is much larger. The producer can vary not only his output, but also his price, the quality of his product, and his selling expenditures. He must take into account not only the reactions of his customers but those of his competitors. If there are few buyers in the market, their behavior also will have a direct bearing on the level of price.

It is necessary under imperfect competition to make a large number of assumptions about the nature of the market and about buyers' and sellers' reactions before one can say what price will be established. While it is possible to reach precise results in a particular case, the number of cases—*i.e.*, the number of possible combinations of assumptions—is very large. One cannot hope, therefore, to develop any satisfactory general formula. At most one

may be able to analyze a few cases which are believed to occur rather frequently in practice.

The discussion in this chapter has by no means explored the full complexity of the problem. Even to describe the possible types of market would involve some such classification system as the following.

Number of sellers	Number of buyers	Type of product
One	One	—
Few	Few	Homogeneous
Many	Many	Differentiated

Combinations of these factors alone—and they are by no means the only ones which should be considered—would yield eighteen major types of market. In justification of the simplified treatment presented here, it can be said, first, that a full discussion would run far beyond the limits of an elementary treatise, and second, that the two types of market discussed in this chapter are among those likely to be of greatest importance in practice.

We first discussed the operation of a market with many sellers, many buyers, and a differentiated product. The results of such a market were found to differ relatively little from those of a purely competitive market. Unless new producers can be excluded, profits will not be any larger than they would be under pure competition. The average price prevailing in the market will be higher than under pure competition, but this is presumably offset by the satisfaction which consumers derive from being able to gratify their preference for their favorite brands or favorite stores. Apart from this, the prices charged by different sellers will differ somewhat, and their profits will also differ, as a result of differences in their individual demand and cost situations.

The second type of market discussed in this chapter was that characterized by few sellers, many buyers, and a homogeneous product. In this case, depending on the relations among the sellers, the price may be highly profitable, may barely cover costs, or may —in the short run—even fall below average total cost. Maintenance of a high level of profits over an extended period of time requires a favorable demand situation, a high degree of coöperation among the sellers, effective barriers to the entrance of new competitors, and avoidance of excessive selling expenditures. It is probable that, on the average, industries containing few firms show higher levels

of prices and profits than would exist under pure competition. The situation varies greatly from industry to industry, however, and even from time to time within the same industry.

Monopoly: Definition and examples. When the fewness of sellers that characterizes one type of imperfect competition, sometimes called oligopoly, reaches the limit—*i.e.,* only one seller—we have a state of monopoly. Monopoly is also attained when the independence of sellers vanishes and is replaced by complete agreement or combination. Monopoly would also be reached by going to the limit of individuality of product. If any seller's product became so completely individualized as to be unique, then this seller would have the monopoly of its sale. Monopoly is defined as the *exclusive control of the supply of an economic good.*

There are monopolies controlling the sources of certain products of nature, such as the Carlsbad mineral waters, the diamonds of Africa, and the anthracite coal of Pennsylvania. The possessor of any unique nonreproducible article, such as a famous work of art, is of course in a monopoly position. A famous actor or musician has a monopoly in the personal services which the public demands of him alone. The United States Government has a monopoly of the service of carrying letters. Patents and copyright give monopolies to inventors and authors, and city governments grant monopolies to corporations in the form of exclusive franchises to operate street railways, to furnish water, gas, electricity, etc. Finally there are certain great corporations and combinations, in fields in which the advantages of large-scale production and combination are especially strong, which have come to positions at least approaching monopoly in their particular fields.

Monopoly price. The position of the monopolist is unlike that of the seller in a purely competitive market, in that the demand for his product is not perfectly elastic. It is represented, not by a horizontal line, but by a downward-sloping curve. In this respect, as the reader will recall, monopoly is like imperfect competition. The seller is able by his own action to influence the price of his product.

Monopoly differs from imperfect competition, on the other hand, in that the monopolist is free to make his decisions and take action without regard to the reactions of other sellers. He has no competitors. His price problem is correspondingly simplified.

The principles which determine monopoly price are those which

govern the first step in the determination of price under imperfect competition. Given a certain demand for the monopolist's product and certain cost conditions, the monopolist will maximize his profit, or minimize his loss, by producing and selling that output at which marginal cost is equal to marginal revenue. However, he does not have to look further and take account of the changes in the demand for his product which may result from this price policy, as does the seller under imperfect competition.

Monopoly and imperfect competition. Our investigation of imperfect competition readily suggests that actually the monopolist differs in degree, rather than in kind, from the ordinary industrialist. The local electric light company may be a monopolist because no one else may sell electric current in the same district; but there is a demand curve for the product of the local utility, just as there is a demand curve for the output of the Ford Motor Company.

Furthermore, recent experience would appear to indicate that the demand for electric current is highly elastic. Therefore the president of the local utility has to make a decision of precisely the same kind as that which Ford must make. The most difficult question which each of these sellers must answer concerns the increase in sales which will result from any given reduction in the price of what he sells. Each must decide what price to set, or what output to throw on the market, with reference to exactly the same criteria.

The essential distinction between a monopolist and a businessman selling in an imperfectly competitive market is that the monopolist sells a product which has no close substitute; whereas the ordinary competitive businessman, although he may sell a highly individualized product produced by no one else, nevertheless must as a rule meet the competition of close substitutes.

For instance, although there is only one manufacturer of the Ford car—the Ford Motor Company—there are other cars which are nearly its equivalent in quality and in price. The availability of these substitutes limits Ford's power with regard to the setting of price. On the other hand, the local water company does not have to compete with any close equivalent and may thus properly be called a monopoly.

From this distinction we may infer that the demand curve for the monopolist's product is usually less elastic than that facing the

businessman who has competitors but sells in an imperfectly competitive market. If the power company increases its rates, its consumers generally have no alternative but to economize. If the Ford Company raises its prices, customers may turn to Plymouth and Chevrolet. Though still only a difference in degree, this is a real and significant distinction between the two for the purposes of economic analysis.

Monopoly and the public interest. The seriousness of a monopoly situation is mostly dependent upon whether the monopoly-controlled commodity is a luxury or a necessity. We have already had occasion to note that the demand for necessities is usually very inelastic. If therefore the necessity is produced or sold by a monopoly, it will pay the monopolist to set his price high, because marginal revenue will be much lower than price and thus, when the monopolist has equated marginal cost and marginal revenue, price will be high. On the other hand, monopoly control of a luxury is no great menace. The monopoly will, in its own interest, generally refrain from charging an exorbitant price. And if its price is excessive, the people can do without this particular article.

Anticipating our later study of the subject of monopoly, it may be stated here that in general the public interest is better served when production is carried on under the regime of competition. There are certain exceptions, some self-evident, others not so obvious. Personal monopolies, such as those enjoyed by the popular singer or actor, are inevitable and can do the public little harm. The monopoly granted by the patent or copyright is generally in the public interest. And there are certain lines of industry, such as railroading, telephone, telegraph, the furnishing of water, gas, electricity, etc., which from their economic nature appear destined to be monopolies, whether or no. In such cases it is futile to put our trust in competition which cannot endure; monopoly must be accepted and made the best of.

Public opinion today generally holds that abuse of monopoly power, particularly in the necessities of life, should so far as possible be prevented, in order to safeguard the public interest, and that, when monopoly is inevitable, government regulation should be relied upon to ensure to the public good service and reasonable prices. For this general opinion there is firm foundation in economic principles.

EXERCISES

1. Assume the following conditions of demand and cost of production for a low-priced automobile:

Price	Sales	Marginal revenue	Average total cost	Marginal cost
$875	100,000	$875	$1,310	$310
780	200,000	685	813	315
700	300,000	540	650	325
635	400,000	440	573	340
584	500,000	380	530	360
540	600,000	320	508	400
500	700,000	260	500	450
465	800,000	220	503	520
435	900,000	195	516	620
410	1,000,000	185	542	780

 (*a*) Plot on a sheet of graph paper the four curves of demand, marginal revenue, average total cost, and marginal cost.
 (*b*) What is the best output for the manufacturer to produce? Reading from the chart, at what price can this output be sold? What is the average total cost of producing this output?
 (*c*) What is the profit per unit and the total profit under these conditions?

2. A new manufacturer of automobiles enters the low-priced field. The additional competition reduces the demand for the product of the first manufacturer, described in Exercise 1. Assume that the new condition of demand for his product is as follows:

Price	Sales	Marginal revenue
$785	100,000	$785
700	200,000	615
630	300,000	490
572	400,000	398
524	500,000	332
485	600,000	290
450	700,000	240
418	800,000	194
390	900,000	166
365	1,000,000	140

 (*a*) Using the chart constructed in Exercise 1, plot the new demand curve and the new marginal revenue curve.
 (*b*) What is now the best output for this manufacturer to produce? Reading from the chart, at what price can this output be sold? What is the average total cost of producing this output?
 (*c*) What is the effect of the appearance of a new competitor on the

price of the product of the first manufacturer and on the volume of his profits?

3. The Federal Trade Commission, a government agency charged with the suppression of monopoly and the preservation of competition, has devoted considerable effort to the prevention of "brand imitation." Does an increase in the similarity or "closeness" of available substitutes tend to make competition more or less perfect? Explain.

4. (*a*) The Harvard Stadium has a seating capacity of about 60,000. The Athletic Committee sets a price of $2 per ticket for the game with Shady Rock and the stadium is half filled. Draw a diagram to illustrate this situation. Reading from your diagram, at what price could the stadium have been sold out?

 (*b*) The Committee sets a price of $4 for the Yale game and is unable to fill all orders for tickets. Draw a diagram to illustrate. Reading from your diagram, how many tickets could have been sold at $4 if the seating capacity of the stadium had been large enough? What price could the Committee have charged and still have sold out the stadium?

CHAPTER XIII

Interrelationship of Prices. Price Theory in Practice

Introduction. We have been investigating the way in which prices are determined by the forces of demand on the buyers' side of the market and supply and cost of production on the sellers' side, under various conditions of pure and imperfect competition and monopoly. It has been recognized throughout this investigation that these immediate forces which directly determine price are themselves the resultants of a great variety of forces. Among these forces, acting thus indirectly upon the price of any good in any market, will always be found the prices of other goods, or the prices of the same good in other markets. The limits of this book do not permit full investigation of the very many price relationships here suggested. But we may profitably give special attention to some of the more direct and fundamental ways in which prices are related to each other.

The present value of future income: Capital and income. The theory of price is incomplete without inclusion of the most fundamental of all price relationships; namely, that between the value of future income and the present value of capital goods. We have learned that all production requires capital.[1] To obtain income man must have capital goods, and it is the production of income which alone gives significance to capital. In a primitive society this effort to secure capital takes the direct form of devising and constructing the weapons, tools, buildings, and other instruments by means of which the productivity of man's labor is multiplied and his enjoyable income increased. Under such simple conditions therefore it is customary for the same man both to bear the final

[1] In the present discussion we shall use the term "capital" in the broadest of the several meanings which have become inseparably attached to it; *i.e.*, as a synonym for wealth, thus including land as well as man-made instruments. The term is generally so used in discussions of investment and when it is desired to give prominence to the distinction between wealth, the fund, and income, the flow. Cf. Chapter I and the footnote on page 47 (Chapter III).

costs of capital—the labor of making the instruments and of keeping them in repair—and to consume its benefits.

Neither capital nor income under such conditions is frequently bought and sold; hence neither acquires an exchange value, and there is no tendency to dissociate capital from income and regard it as an end in itself. Capital is viewed in its proper light, as a means toward an end. Like other means, each unit of capital is more or less prized according to its efficiency in achieving the end in view; that is, its power to produce income for its owner.

In the social order of today there are to be found many instances of this direct and visible relation between capital and income. Men sometimes gather materials and with their own labor construct some simple capital instrument whose benefits will accrue to themselves directly. The farmer may cut and shape a handle for his ax, or build a fence or wagon shed. A blacksmith may make some of his own tools. In such cases, the relation of capital to income is clearly grasped; one sees that it is the expected income which bestows on capital its importance, and that the owner's appraisal of the capital is but a reflection of his evaluation of the benefit which he hopes to obtain from it.

Property and capital. But these simple cases are not typical of present-day society. So complex has our economic structure become that we often lose sight of the fact that capital has no significance when divorced from the benefits which it produces. Chief among the attributes of our social system which have this effect is the fact that the ownership of capital has been subdivided into a large variety of property rights, which are bought and sold in great numbers.

Those property rights are usually represented by documents of different kinds—stocks, bonds, mortgages, contracts, and the like—which frequently have stamped upon them some nominal valuation (for example the $100 par value on a stock certificate), which misleads the owner into the belief that these documents have a "normal" or "true" value in themselves. The development of many agencies with the purpose of facilitating trading in these property rights—such as banks, brokerage and investment houses, stock exchanges, and so on—has aided in beclouding the true relation of capital to income.

As we have learned, property is but a right to the future benefits of wealth (capital). The benefits are produced by the capital which

underlies the property; it is the function of the property right to divide this income among the owners of the capital. It is apparent therefore that the principal reason for the existence of the certificate of property is that it affords a convenient method of transferring ownership in capital from one man to another and of subdividing the ownership. These certificates are bought and sold as representatives of the valuable thing (capital) in which they transfer ownership; hence the action of the buyers and sellers in the market will result in giving the certificates the same value that, in the opinion of these buyers and sellers, should be placed on the capital which they represent.

If for example a manufacturing concern has issued one thousand shares of stock and these shares sell in the market for $150 each, there can be but one reason for this; namely, that in the opinion of investors a one-thousandth share of the owners' equity in this corporation is worth $150. The fact that the stock has stamped upon it a nominal value of $100 or $200 or $50 or any other figure is not competent to determine the value of the share.

Present worth. The real problem therefore is to discover how the value of capital goods is determined. Here we must repeat that capital is desired only because of the income it produces and is bought only as a means of obtaining this income. Let us suppose that a given machine will last ten years and produce a net income of $100 a year. If this machine is offered for sale to the owner of a factory, he will consider that by owning it he can increase his income during the next ten years by the amount of $100 a year, or $1,000 in all. Whatever price he is willing to pay for the machine will be a measure of the present value he puts upon this thousand dollars, distributed as it will be through the ten years to come.

To change the illustration, we may suppose that an acre of land is offered for sale to a man who has reason to believe that it can be made to produce a net income worth $10 a year forever. In this case too, the price offered by the buyer for the land will be his estimate of the *present worth* of an annual income of $10 which continues indefinitely.

Or again an investor may be offered a bond which will be redeemed for $100 at the end of five years and in the meantime pay $5 a year interest. Such a bond would be viewed as an opportunity to receive a series of money payments during the next five years: $5 one year from now, another $5 two years from now, another

$5 three years from now, $5 more four years from now, and a final payment of $105 after five years have elapsed. The amount of money that the investor is willing to part with today, in order to buy the bond, must be an indication of what he considers this series of future payments to be worth at the present moment.

This is the essential thing in calculating the value of any capital good or of any property right in capital goods. It amounts to placing a present value on an expected future income of a certain size and form and duration.

Present goods more desirable than future goods. It is a well-known fact that men do not consider a sum of money which will accrue in the future to be worth its face value now. The investor referred to in the preceding illustration will not calculate the present value of the bond by simply adding together the different payments which will accrue in the future to its owner. Such a calculation would give the bond a present value of $125, and the buyer paying this price for it would eventually get back just the amount of money that he paid, though he would have to wait five years for most of it.

If one offered for sale a promissory note which would pay its owner $100 twelve months from today, he would find no one willing to pay $100 for it now. Nor would this reluctance to buy this future income at its full face value in present money be due only to risk of loss. If the $100 were locked up so safely that all doubt regarding its security were removed, still no one would be willing to pay out $100 now and be content to get it back after twelve months had passed.

In the mind of the ordinary man the present value of the future dollar is less than that of today's dollar; hence sums of money which accrue in the future will be bought only if they can be had for less than their face value. And the price that will be paid for capital which gives its owner the right to receive income in the future will be lower, by some substantial amount, than the value of this future income when it accrues.

Interest. If asked why he will not pay $100 now for the right to receive $100 one year hence, the practical man will probably answer that he is unwilling to lose interest on his money. Because of the existence of interest, it lies in the power of any man who is willing to exchange present wealth for future wealth to make the exchange on terms which will bring him a future return of larger

value than his present payment; and as long as this opportunity to gain interest on the money they lend is open to all men, no one will consent to pay its full future value in present money when buying income.

Interest may be defined as *the premium which present goods command over similar goods due in the future.* The man who lends $100 for one year at 5 per cent interest exchanges $100 worth of present goods for $105 worth of next year's goods, the premium of $\frac{5}{100}$ being required to give the future goods a present value equal to that of the present goods which the lender surrenders. As the interest rate rises and falls the premium which must be added to make these two amounts—one present, the other future—of equal present value rises and falls accordingly. It is therefore the *rate of interest* which determines how much larger than the present sacrifice the future return must be, in order to cause the exchange of these two to occur.

The link between future and present values. We need not at this point attempt to explain why interest exists or how the rate of interest is determined; these are problems to be solved later in our study. For our present purpose, it suffices to accept the existence of interest as a feature of the present business world. Our immediate concern is to show that the rate of interest supplies the link between future and present values which enables us to translate the one into the other.

Given the rate of interest, it is a simple matter to calculate the equivalent future value of any amount of present money by means of the following formula:

$$A_n = P(1 + r)^n, \text{ in which:}$$

A_n represents the future value, any number of years hence;
P represents the present value, or principal;
r represents the rate of interest;
n represents the number of years.[1]

[1] The demonstration of this formula is as follows:

At the end of the first year, the amount (A_1) is the sum of the present value, or principal, and one year's interest; that is,

$$A_1 = P + Pr = P(1 + r).$$

That is, the amount at the end of the first year is the principal multiplied by $(1 + r)$.

This amount, $P(1 + r)$, is the new principal for the second year. The amount

For example, when the interest rate is 5 per cent, \$100 invested now will be worth \$105; *i.e.*, \$100 $\times$ 1.05, one year from now; \$110.25; *i.e.*, \$100 $\times (1.05)^2$, two years from now; \$115.76; *i.e.*, \$100 $\times (1.05)^3$, three years from now; and so on.

This same formula may be adapted to the reverse problem of determining the present worth of a given future value by merely rearranging its terms, as follows:

$$P = \frac{A_n}{(1+i)^n},$$

the symbols having the same meaning as in the preceding formula. To illustrate, assuming again an interest rate of 5 per cent, \$100 due one year hence is worth now \$95.23; *i.e.*, \$100 $\div$ 1.05; if due two years hence, it is worth now \$90.70; *i.e.*, \$100 $\div (1.05)^2$; if due three years hence, it is worth now \$86.31; *i.e.*, \$100 $\div (1.05)^3$; and so on.

These numerical examples assume that the interest rate is so much per annum and the interest is *compounded* annually. The formula holds, however, for any unit of time. All that is required is that the rate of interest be stated with respect to the same period as is used in compounding. For example, it is required to find the amount of a loan of \$5,000 for two years at the rate of 1 per cent a month, interest compounded semiannually. Since the interest is compounded every half year, we must take a half year as the unit of time. Then we must call the rate of interest, instead of 1 per cent per month, 6 per cent per half year. Instead of saying the loan is for two years, we must say it is for four half years. Then our formula becomes $A = 5{,}000(1.06)^4$; and when we solve, we find that $A =$ \$6,312.38 $+$.

The formula which determines the present worth of a given future value will serve likewise to place a present worth upon a series of future values. What is the present worth of a series consisting of \$200 due one year hence, \$500 due a year later, and \$100 due at the end of the third year, if the rate of interest is 5 per cent?

at the end of the second year (A_2) will be found in exactly the same way as for the first year; that is, by multiplying the principal at the beginning of the year by the factor $(1 + r)$. Thus,

$$A_2 = P(1+r)(1+r)$$
$$= P(1+r)^2.$$

In the same way, we find at the end of the third year the amount is $P(1+r)^3$, at the end of the fourth year, $P(1+r)^4$, and so on for any number of years.

By means of the formula $P = A_n/(1 + r)^n$, we find the present worth of these future amounts to be respectively \$190.48, \$453.52, and \$86.31. The total of these three amounts, \$730.31, is the present worth of the series of future values. By this process, one evaluates the ordinary bond. Thus a 5 per cent \$1,000 bond due in seven years is valued as a series of six annual payments of \$50 and a final payment of \$1,050 at the end of the seventh year.

It is frequently important to evaluate a uniform perpetual series of future values, as in the case of a perpetual government bond. The general formula for the present worth of a uniform perpetual income is $P = \frac{i}{r}$, in which P stands for the present worth, i stands for the periodical payment, and r stands for the rate of interest.[1] Thus, assuming an interest rate of 3 per cent, the present worth of a 4 per cent \$1,000 perpetual bond (like the British consols) would be $\$40 \div {}^{3}/_{100} = \$1{,}333.33$.

The discount process. Although it is possible, by means of the interest rate, to calculate either future values from present or present values from future, and although in the business world there is continual use for both of these calculations, it is the latter process which has the greater importance. This is true for the reason stated at the beginning of this chapter; namely, that the economic activity of mankind is directed toward the obtaining of income, and all income which has power to give value to capital lies in the future.

The valuation of capital instruments and property rights is incidental to this struggle for income, being a product of the in-

[1] This formula is derived, as a special case, from the general interest formula, remembering that, in this case, the interest is not compounded but paid each year, so that the principal is brought back each year to the original amount. According to the general formula,

$$P = \frac{A_n}{(1 + r)^n}.$$

Assuming a single interest period, $n = 1$. Then,

$$P = \frac{A}{1 + r}.$$

But, $A = P + i$.

Hence, $P = \frac{P + i}{1 + r}$;

$P + Pr = P + i$;

$Pr = i$;

$P = \frac{i}{r}$.

cessant buying and selling of the means by which income is produced. Hence the calculation of the present values of future incomes of every conceivable variety as regards size and duration is one of the commonest events in the business world. The name given to this calculation of present values of future incomes is the *discount process.*

Summarizing the preceding discussion, we may say that this process requires these two fundamentals: (1) a knowledge of the probable future value of the income, (2) a knowledge of the interest rate. The first of these fundamentals is the basis of the calculation—the future value whose present worth it is desired to obtain. The second supplies the mechanism for determining the present worth; the process being to divide the future value by one plus the rate of interest raised to a power equal to the number of years which must elapse before the future income accrues.

Capital and income. The relation of capital to income may be looked at from either of two points of view; namely, that of the physical relationship and that of the value relationship. From the physical point of view, it is capital which produces income; an orchard produces a certain number of barrels of apples; a factory, so many pairs of shoes or other units of output; a house, so many years' shelter; and so on.

But the value relationship is the opposite of this. The value of the capital does not produce the value of the income. The value of the apples gathered from an orchard will be established in the market in which the apples are sold after they have been produced, in accordance with the laws of market price as developed in preceding chapters. Setting a value upon the capital will not avail to dictate the terms upon which its product can be sold. On the other hand, a given set of capital instruments may be very productive, in the sense that it turns out a large stream of products, but if these products cannot be sold at profitable prices, the value of the capital instruments will be very low, or even nonexistent. This is clearly evident of capital when the demand for its products has declined or disappeared. Abandoned farms, factories, and railroads bear testimony to the fact that the value of capital goods is derived from that of their income.

Discount and the laws of price. It is essential that we understand exactly how this process of discount is related to the general laws of price which we have previously developed. We have found that,

under conditions of pure competition, prices are in general determined by demand and supply and that, under imperfect competition, prices are determined by decisions of the sellers in conjunction with the conditions of demand and with consideration of actual and potential costs. We have learned that back of demand is marginal utility, and back of supply and sellers' actions, cost of production. Now it appears that values are determined by discounting future expected incomes. Are we faced with two conflicting principles?

To make our question concrete, suppose we ask: Why does a particular tract of land sell for $200? Is it because this is the price determined by the forces of demand and supply in the market, or is it because this is the present worth of an expected perpetual income, say $10 a year discounted at 5 per cent? The answer is yes to both questions, and the absence of any conflict will be apparent as soon as we inquire into the part which discounting plays in the market.

Values are not determined directly by discounting. In the simple numerical examples which we have thus far employed, we assumed certain future values and we assumed a certain rate of interest. In the practical world, neither of these essential factors is objectively determined. The future incomes—just because they are future—are uncertain, more or less. The rate of interest is the rate at which any given person will discount the future.

Every potential buyer and every potential seller must come to his own decision as to what these factors are in any given case. A, on the basis of the information available to him and his best judgment, decides that the tract of land in question will probably bring forth such and such future incomes, that these incomes will have certain values to him, that he is prepared to discount the future at a certain rate of interest; with those factors thus estimated, he goes through the discounting process and determines how much he is willing to pay for this piece of land. B makes similar estimates, arriving possibly at different figures from A at every point, and finally emerging with the price he is willing to pay. So of all other possible purchasers. And the present owner of the land must himself go through a similar process in order to determine at what price he will be ready to sell.

When the potential buyers and sellers meet in the market, the discounting process has already taken place in the mind of each.

If the tract of land in our example sells for $200, it is because the seller considered it worth something less than that to him, and the purchaser considered it worth something more. In arriving at his judgment, each employed the discounting process. The discounting process is itself one of the forces that determine demand and supply, which in turn determine price. Discount, so far from being in conflict, is itself included in the principle of demand and supply.

We recognize of course, as in all such theoretical analyses, that every person does not consciously go through the mental process here visualized. Some buying and selling, as we have seen, is more or less irrational. But back of demand and supply will always be found human judgment discounting expected future incomes.

Even where—as appears to be the normal case on the speculative stock and produce markets—dealers seem to make their decisions, not with any regard to discounted future incomes, but on the basis of expectations as to what prices will be in the future, the fundamental principle is still in operation, though somewhat obscured. For in the last analysis, back of all estimates of future prices, there is always someone's forecast of future income and its discount at a rate of interest. The dealers are simply cutting corners to arrive at the result. The market rate of interest itself is a composite judgment of the relation between future and present values.

Discount and the cost of production. Since many capital instruments are produced in a succession of separate standardized units in anticipation of an income to be derived from them at a more or less remote future date, the question arises as to how the producers of such instruments are able to relate their present costs of production to a demand which is dependent upon a long deferred income. Such instruments are staples in the market. They are constantly being reproduced as a matter of routine business practice; day after day they appear in a steady stream upon the markets of the world. The producers of these instruments, after they have offered their supplies for sale, must be content with a price determined by the laws of market price, in the operation of which the discount process will govern the demand of the buyers. But what if the estimate thus placed by the buyers upon any given quantity of the capital instruments differs from the cost of producing this quantity?

The discrepancy will be resolved, as we have learned, through the power which the producers have, either to avoid future costs

by giving up or diminishing the production of these instruments, or, on the other hand, to increase future costs by increasing their output.

Their decisions on this point will, after an interval of time, exert an influence upon the value of the capital instruments they are producing. If they produce an increased quantity of them, the services rendered by them will increase in quantity and command a lower price; hence the discounted value of the capital instruments will fall. On the other hand, if the producers reduce their output of the capital instruments, the peculiar benefits which flow from the instruments will become scarcer and their price will rise, hence raising the discounted value of the capital. In other words, it is the willingness or unwillingness of the producers to bear future costs which, by affecting the value of the income from capital, brings marginal utility and marginal cost into harmony.

In the case of standard and staple capital instruments, the markets for which have settled to a point of relatively stable equilibrium, the producers may confidently take the present price as an indication of what the future price will be. If the present price of the instrument stands very much above the cost of making it, existing producers will be tempted to expand their output, while at the same time, other entrepreneurs will be tempted by the prospect of large profit to engage in its production. As the number of instruments in use increases, the value of the income which one instrument can earn will decrease, and consequently the discounted value of the instrument will fall. The marginal cost of producing the instrument will rise as attempts are made to supply it in larger quantities. As a result the discounted value and the marginal cost, which at the outset were different from each other, will tend to draw together.

On the other hand, if the present price of the capital instrument were below the marginal cost, entrepreneurs desirous of avoiding the losses caused by this situation would abandon the production of the instrument entirely or contract their output of it. This restriction in the supply of the instrument would cause the income it produces to grow scarcer and higher in price, with the result that the discounted value of the instrument would rise, and the diminution in the production of the capital instrument would reduce the marginal cost of production. Again there would be a tendency for

the two—marginal cost and discounted value—to approximate each other.

In the case of capital goods which are not being currently reproduced, discounted value may part company from cost of production indefinitely. Then it is discounted future income, not cost of production, that determines the value. The construction of dwellings, office buildings, hotels, and factories will not be undertaken unless the builders expect the selling price to be as high as the cost of production. But after the costs involved in the erection of a given building have been borne and the building is offered for sale, the price will be determined by the buyers' estimate of the present worth of its future products. A summer resort constructed at great expense in an unpopular place will command a price lower than its cost of production, because the discounted value of the net income to be derived from its operation is low.

Prices of the same good in different markets. Turning now to a different type of price relationship, let us suppose that the price of wheat in Chicago is 80 cents a bushel and in New York $1.05, and further that the entire cost of transporting a bushel of wheat from Chicago to New York is 15 cents. There is evidently a profit to be made by buying wheat in Chicago, shipping it to New York, and there selling it, and there is a class of businessmen ever on the alert to take advantage of such opportunities. Such dealers will immediately enter the market, seeking the profit of 10 cents a bushel which the existing conditions indicate.

But these very operations will inevitably change the conditions. The dealers enter the Chicago market on the demand side, increasing the demand for wheat and so raising its price. In New York, on the other hand, they appear upon the seller's side and tend to increase the supply and lower the price. Even so, the margin of profit, though declining, continues for a time; and a considerable quantity of wheat will be transferred from Chicago to New York.

The prices on the two markets will approach each other until finally the difference is only 15 cents. Let us suppose the Chicago price has risen from 80 cents to 84 cents while the price in New York has fallen from $1.05 to 99 cents. The two prices are now in equilibrium with respect to each other, differing by exactly the cost of transportation. In such manner the prices of the same good in different markets may react upon each other.

Such reaction is obviously possible only where there would other-

wise be a difference in price greater than the cost of transportation from the low-price market to the high-price market. If the price of brick is $10 per thousand in New York and $8 per thousand in Chicago, while the cost of transportation between the two cities is $20 per thousand, these prices are evidently independent of each other, since no one would be tempted to buy brick in one market for sale in the other.

The extent to which prices in a given market are affected by the prices of the same good in other markets thus depends upon the cost of transportation in relation to price. Why were the commodities which entered into the long-distance trade of the medieval merchants so generally articles of luxury—jewels, silks, fine wines, spices, perfumes, etc.? The answer is that heavy transportation costs made it unprofitable to bring to market from remote places anything that could not be sold at a high price. Farm produce, for example, could be sold profitably only in the markets close to the place of production.

Since that time the transportation system has become so extraordinarily efficient and the price of transportation so low that the province of trade has been enormously extended. Cheap and bulky goods, such as wheat and bananas, are shipped thousands of miles over land and sea to be placed at the disposal of distant consumers.

With each step in the progress of lower transportation costs, the realm of these price reactions is widened. For many products, such as wheat, cotton, the precious metals, etc., the markets of the whole world are thus tied together. This subject will be further investigated when we undertake the study of interregional trade, at which point we shall better realize how vitally these price relations affect the well-being of mankind.

Prices of competing goods. In the winter of 1925-26 there was a long-continued strike of the miners in the anthracite coal mines of Pennsylvania, which for a considerable time virtually stopped the flow of such coal to the markets. The resulting decline of supply greatly raised the price of anthracite. As a consequence, people sought other kinds of fuel, such as bituminous coal, coke, etc. This meant an increased demand for such fuels and in consequence higher prices.

Suppose there is in a certain market a sudden increase in the supply of beef. Beef prices will decline in consequence, and more people than usual will buy beef. This will entail a decrease in the

demand for lamb, let us say, and the price of lamb will therefore go down.

These events illustrate another sort of price relationship; *i.e.*, the relation between the prices of substitutes or competing goods, with which the reader has already become acquainted through earlier chapters. We have the general principle that the prices of competing or substitute goods are so related that they tend to rise and fall together.

Prices of complementary goods. Suppose some change in conditions of production should bring about a material increase in the price of bread. People would naturally incline to diminish their consumption of bread, seeking to find similar nutritive satisfaction from other kinds of food. The result might be, in accordance with the law of competing prices, a rise in the price of meat or vegetables. But what, if any, would be the effect upon the price of butter? Butter is chiefly used in connection with bread. Using less bread, people would want less butter. The demand for butter would decline and its price would fall. Similarly a fall in the price of bread, by causing its increased use, would increase the demand for butter and tend to raise its price.

This is one example of a pair of complementary goods, meaning goods which are used only or chiefly in conjunction with each other. There are plenty of other examples. In the extreme case, the use of the one good is possible only in connection with the other; for example, tennis rackets and tennis balls are completely complementary goods. It is evident that the prices of complementary goods tend to move inversely to each other, the reverse of the relation between the prices of competing goods. For example, horses and mules are substitute or competing goods; their prices tend to rise or fall in unison. Horses and wagons are complementary goods; their prices tend to move inversely.

Prices of competing cost goods. The price relations of competing goods and complementary goods arise out of the reactions of the buyers; they are on the demand side of the market. There is another relation between competing goods (substitutes), which we have not as yet encountered. Labor is normally devoted to those pursuits which offer the greatest return. Consider the case of a small town in which there is one important manufacturing industry making, let us say, grindstones. A rise in the price of the product of this enterprise will cause its owner to seek to expand his output;

for this purpose he will require more labor, which he will draw, by offer of higher wages, from the dairy farms and market gardens of the neighborhood. The resulting shortage of agricultural labor will cause a decreased supply of milk and vegetables in the local market, with a consequent rise in their prices. Thus a rise in the price of grindstones has caused a rise in the price of vegetables.

Products such as these, both of which require the same cost element, make competing demands for the factors of production—competing in this example for the use of labor. They may be called *competing cost goods.* Their prices evidently tend to move up and down together. This relation is quite distinct from the one which arises from the demand side; that is, from the fact that a rise in the price of one good increases the demand for all substitutes.

Prices of joint cost goods. Where cattle are slaughtered for beef, hides are made available, as well as horns, tallow, and a great number of other products. These products appear jointly as the results of the same production costs. Having determined to produce a certain quantity of one, the industry will inevitably bring forth the corresponding quantities of the other products.

Where the products are of unequal importance, the less important is called the *by-product.* For example, the growing of cotton necessarily yields a crop of cotton seed. The latter has its uses, but in general cotton would not be grown for the sake of the seed; that appears incidental to the production of the primary crop.

Such products as these are said to be the results of joint costs, and their prices have a definite relation to each other. Let us suppose that something causes the price of cotton to be especially high. To take advantage of this, the planters will increase their production of cotton, thereby inevitably increasing the supply of cotton seed. The price of the latter will consequently be reduced. A fall in the price of cotton would discourage production, thus leading to a decline in the supply of cotton seed and a rise in its price. The prices of joint cost products thus tend to move inversely, being the same relation as was found to exist between the prices of complementary goods. In recognition of this analogy, *joint cost goods* are sometimes called *complementary on the supply side.*

Prices of "tandem goods." Where one good is the material from which another good is made, there exists still another sort of price relation. If the price of wool rises, there is caused thereby an in-

crease in the marginal cost of producing woollen cloth. This will cause a reduction in the supply of woollen cloth; *i.e.,* at any given price the quantity offered for sale will be less than before. Hence the price of woollen cloth rises. This is a simple example of a principle that is a commonplace of everyday observation; *i.e.,* that the prices of manufactured goods tend to vary directly with the prices of the materials that enter into their production.

Price theory in practice: Practical aspects of the theory of value. The very considerable time which we have devoted to study of the theory of price is justified by the importance of these concepts in every department of business and practical affairs. Inasmuch as the social system of the United States rests upon private ownership of wealth and free exchange, problems of value are at the heart of the practical enterprises of the people.

This appears nowhere more clearly than in the field of investment. The ownership of corporations (in the broadest sense) is expressed in their stocks and bonds. These securities are bought and sold, in some cases daily or even hourly on the exchanges, in other cases more or less rarely and privately. Values are determined thus by the action of demand and supply, and back of demand and supply is always the process of discount or capitalization of income. Consciously or unconsciously, every buyer or seller or owner of corporate securities is making practical use of value theory in reaching his decisions to buy or to sell or to hold.

That the values of corporate shares sometimes appear to be quite out of harmony with the current dividends, so far from being denial of the validity of our theory of value, is the best proof of its truth. Before the sensational stock market crash in the United States in the autumn of 1929, the prices of American stocks generally were very high. For example the common stock of the United States Steel Corporation reached its high point of 261¾ on September 3, 1929. The annual dividend rate being 7 per cent, purchasers at this price and owners who held at this price were apparently satisfied with a return of only 2⅔ per cent.

Now the fact is that American stockholders are not satisfied with any such return as that. The explanation is that these investors were not so naïve as to discount only present income; they were acting in accordance with sound theory and discounting the future incomes expected to flow from shares in the United States Steel Corporation, based upon surpluses already accumulated and ex-

pected future earnings. September, 1929, witnessed the culmination of a long period of buoyant optimism among investors, and there was a general anticipation of high future earnings for American corporations. The reaction was seen in high values and low apparent yields all along the line.

That these rosy views of the business future were exaggerated and not warranted by the facts was the opinion of some of the more conservative observers as early as the summer of 1929, and the correctness of this opinion was demonstrated by the collapse of values that occurred in October and November. United States Steel dropped from its high price of 261¾ (on September 3) to 150 on November 13, at which price, it stood to yield on the basis of the past year's dividends 4⅔ per cent.

The important lesson for our present purpose is that it was the investors' and speculators' estimates of future facts, not their theory of value, that was wrong. And this brings out clearly the principle that all value is the resultant of human judgment, human estimates of future events. Obviously, no one is sure of the future.

In any broad market, such as the stock market, there are dealers of all degrees of knowledge and judgment, from those who, by resort to all possible sources, seek to inform themselves of conditions and exercise the best judgment that is humanly possible, to those who act, without knowledge or intelligence, on the basis of rumors or "tips" or mere guesswork. The resultant demand and supply are the consequences of the composite judgment of these buyers and sellers, weighted in proportion to their respective resources. Knowledge and ignorance, sound judgment and recklessness, conservative investment and pure gambling, all combine to determine prices, through the chain of cause and effect which we know as the laws of price or value.

The overworked term "value." Although the concept of value as used in the science of economics is thus basic in practical affairs and is so recognized by practical men, there have developed in certain administrative fields other concepts, closely related to value, though not precisely the same, which generally go by the name of "value."

Thus the accountant finds it necessary to assign values to the various items in his statements. Starting with the basic notion of value, accounting practice has evolved certain concepts which serve its purposes but depart more or less from the economic notion of

value. Thus items of wealth, such as land, buildings, machinery, etc., may be carried on the books on various bases of cost, cost less depreciation, replacement cost less depreciation, etc., even though these "values" may be known to be higher or lower than the considerations for which the articles could be sold. "Goods in process" of manufacture are "valued" on the cost accountant's books on such bases as cost of materials plus direct labor cost, with or without a share of overhead, even though in certain cases there may be no possibility of sale of such partly finished products except to another manufacturer in the same line; in the absence of such a buyer, these goods may have no real value except as scrap. In such cases as these, the term value is frankly used to mean something different from the economic concept of value. It is always unfortunate when the same term is used with diverse meanings, but a clear understanding of the usages employed will generally suffice to avoid confusion.

The actual value of a corporation to its shareholders may generally be best shown by the total market value of its stock. This may, however, differ widely from the net value of its assets as shown by the balance sheet.

The discrepancy will be due to a variety of causes. In the first place, the accountants' valuations may depart from market value, as already seen and, such as they are, they may not be kept strictly up to date. Secondly, the balance sheet may not include certain property rights of an especially intangible or elusive nature, such as the character of management, certain franchise or contract rights, the "good will" of an established clientele, etc.—factors which are combined under the title of "going concern value" and upon which the future earnings of the enterprise may in large part depend.

The balance sheet valuation, though serving effectively its own purpose, may thus depart materially from the economic concept of value. That the former is not the real value is demonstrated by the very fact that buyers and sellers of the corporation's securities arrive at values different from those shown by the balance sheet.

In the practical task of regulating the rates of public utility corporations, such as the railroads, the telephone and telegraph companies, gas and electric companies, etc., it is found necessary to ascertain "values" of the capital invested, in order that the legally determined rates may permit a "fair return" on the investment. In this task the various official regulatory bodies, such as the United States Interstate Commerce Commission, the state railroad and

public utility commissions, etc., and the courts, have developed a technique of "valuation," which scarcely pretends any longer to seek to arrive at the economic concept of what the property would sell for. What is actually sought is a "rate base" which shall effectively serve the purposes of regulation. That such a rate base continues to be called "value" should not confuse the well-informed.[1]

A modified concept of value is likewise used in taxation, especially in America, but unfortunately with less candor than prevails among those who deal with public utility rate making. Examinations of the property tax laws of the forty-eight states would show that these taxes are imposed universally upon the value of the taxable property and that, with few if any exceptions, the lawmakers had in mind the concept of value as used in economic science. The legislatures have gone out of their way to specify "market" value, "cash" value, "actual," "fair," "true and just" value, what the property would bring if sold "by a willing seller to a willing buyer," etc.—terms which indicate their adherence to the economic concept of value.

Property tax administration, however, has not always succeeded in carrying out this idea of value, though still generally maintaining the pretense of a true valuation. Assessed values are very often not the same as true values, and they vary all the way from considerably above the true value to as little as 5 to 10 per cent of it; as a rule assessments are below true value. It is safe to say that tax administration will never be competent to place true values upon all the complex kinds of property which in America are taxable. For certain classes of property it may become necessary to emulate the frankness of the rate-making authorities and set up a more or less arbitrary "tax base" which will be workable.

The use of the term value in these varying meanings has been the source of much confusion—confusion which is to be escaped only by recognition that book values, rate-base values, tax values, etc., are generally neither the same thing as value in the economic sense nor consistent with each other. It is not, as sometimes assumed, indisputable evidence of fraud or corruption when a railroad reports one "value" to the rate-making commission and an-

[1] See Tunell, G. G., *Value for Taxation and Value for Rate Making.* Columbia: Proceedings of the Twentieth National Tax Conference, 1927, pp. 263–79.

other to the tax assessor. Nor is there the supposed fine logic in the proposal to stop tax evasion by permitting the state to buy in any property at the valuation on the owner's tax return. Discrepancies between market and book values do not necessarily demonstrate either foolish investing or crooked accounting. And finally the situation does not demonstrate hopeless confusion as to the meaning of value.[1]

EXERCISES

1. Suppose $100 is deposited in a savings bank which pays 2 per cent interest on such deposits.
 (*a*) What would this deposit be worth in three years' time if no money were withdrawn in the meantime?
 (*b*) What would it be worth in six years?
 (*c*) What would it be worth in fifteen years?
 Note: Solve (*a*) exactly. For (*b*) and (*c*) work out the formula only.
2. Suppose the interest rate is 4 per cent. What is the present value of a promissory note for $1,000 due in two years?
3. A stand of growing timber will be marketable three years hence at an estimated value, over and above the cost of cutting, of $5,000. What is its present value if the interest rate is 5 per cent?
4. Suppose the interest rate is 3 per cent. What is the present value of a $1,000, 4 per cent bond maturing three years hence? This description means that the owner of the bond will receive $40 per year for three years and a redemption payment of $1,000 at the end of the third year.
5. A certain piece of land can be made to return a perpetual net income of $600 a year. What is its present value *(a)* at a 5 per cent rate of interest? *(b)* at a 4 per cent rate? *(c)* at a 6 per cent rate?
6. In general, if income value remains unchanged what is the effect on the value of capital *(a)* of a rise of the interest rate? *(b)* of a fall of the interest rate?
7. Draw diagrams illustrating the relation between the prices of competing goods.
8. Draw diagrams illustrating the relation between the prices of complementary goods.
9. Draw diagrams illustrating the relation between the prices of joint cost goods.
10. Draw diagrams illustrating the relation between the prices of tandem goods.
11. Between January and November, 1946, the price of the common stock of General Motors Corporation ranged between 80⅜ and 47¾. The Corporation paid dividends on this stock at $3.00 per share

[1] For an interesting discussion of these matters, written from a somewhat different viewpoint, see Edie, L. D., *Economics: Principles and Problems.* New York: Thomas Y. Crowell Company, 1932, especially pages 190–201.

during 1944 and 1945. The book value of the stock on December 31, 1945, according to its published balance sheet, was approximately $26.50 a share. Account for the variations in price and for the apparent "discrepancy" between book value and market value.

CHAPTER XIV

Principles of Money

Nature of money: Some historical examples. The modern economic structure is reared on the foundation of division of labor and exchange of products. Exchange, in its modern omnipresence and complexity, would be impossible without an efficient medium, which brings us to the subject of money, the medium of exchange, and its correlative, banking.

Today as in the past money takes a variety of forms. There are coins of gold, silver, nickel, bronze, and other metals. There are pieces of paper inscribed with promises and other statements. Cattle and sheep have been used as money, as well as wheat, corn, tobacco, musket balls, beaver skins, fish, meat, shell beads, gold dust, and so on. Monetary history discloses that each community adopts as its money some commodity or commodities of common use among the people, something that is desired by most of the people most of the time.

One of the earliest forms of money of which we have knowledge is the sheep and cattle of peoples in the pastoral stage of economic development. Most of the people derived their living from their flocks, and these represented their principal form of wealth. One would always welcome an addition to his flock, and thus it became the natural thing to use the domestic animals as money. The word "pecuniary," derived from the Latin word *pecus* for cattle, is an interesting heritage of the early stage in which cattle and money were in a sense synonymous.

The common characteristic. All these forms of money have one feature in common—general acceptability in exchange for other things. The things used for money acquired this characteristic originally from the fact that they were desirable for their own sake; *i.e.,* they served to satisfy human wants. In certain cases, it is true, the thing used as money could not be applied directly to the satisfaction of wants. An example is the wampum used by the New

England colonists. Though directly useful to the Indians, it satisfied directly no want of the colonists. Yet it became useful as money as soon as the habit of accepting it in exchange became established generally throughout the community.

For the individual, it is the fact that money is generally acceptable which makes him willing to accept it when offered in exchange for his goods or services. If a person who has valuable goods to sell will take in exchange a coin of gold, or a piece of paper engraved with the promissory note of a bank, or a bale of tobacco leaves, it is not ordinarily because he has any need of these particular articles. He accepts them simply because he knows that at any time he can find other people willing to take them in exchange for whatever they may have to sell to him. This is the fundamental characteristic of money; so long as this qualification is met there may be the utmost diversity as to other features.

The general acceptability of money is always more or less local. Most American merchants would refuse to accept a Bank of England note in payment for goods sold, and a Canadian quarter is not generally acceptable in America except close to the border. In general each nation has its own monetary system, and the money of one nation is usually not acceptable within the borders of other nations. By general acceptability we mean acceptability within the particular community.

Definition of money. We have now all the essentials of the scientific definition of money, which was stated but not explained in a previous chapter. *Money is anything which is generally accepted in a certain community in exchange for other things.*

Functions of money. The principal function of money is to facilitate exchange, in other words, to serve as a *medium of exchange.* Goods and services are exchanged for a common medium, money, which can then be used to purchase other goods and services at the buyer's convenience. Without such a medium, exchange can operate only with great difficulty. The maker of shoes, for example, who wished to exchange some of his product for bread could succeed only when he had discovered the coincidence of a baker who wanted to exchange bread for shoes. A similar problem would face the shoemaker in connection with every other good which he sought to obtain in exchange for his own product, shoes. To exchange so conducted, the name of *barter* is applied. The modern economic order, based on division of labor and exchange of prod-

ucts, could never have become established had not the introduction of money released exchange from the restrictions of barter.

A second essential function of money is to furnish a *standard of value.* This function follows necessarily from the first. It is inevitable that the unit of the good which serves as the medium of exchange should be the unit for measuring the values of all exchangeable goods. Thus money furnishes the universal unit by which alone all kinds of wealth and services may be measured and compared and totaled. This function of money is as essential to the existence of the modern economic system as is the presence of a medium of exchange.

Money performs a third function in serving as a *store of value.* In so doing it enables us to separate our sales from our purchases. One can sell what he produces for money and later, according to his convenience, buy the various things he needs; one can even, by borrowing money, buy first and sell later. The consumer "holds" a certain portion of his wealth in the form of money, as does also the producer.

Government control of money: In general. Money originated in each community naturally and unconsciously as the result of slowly established custom. People simply fell into the habit of accepting some one commodity in preference to others in exchange for goods and services which they had to sell, much as boys at a certain stage in their development fall into the habit of using marbles as a medium of exchange.

Thus the monetary system became firmly established in the mores of the community, and eventually the government took cognizance of it and proceeded to define and enforce it. For centuries the control of the monetary system has been jealously guarded by kings and rulers as one of the most precious attributes of sovereignty.

Today the system in all its details is determined by law, and it is this modern legal monetary system with which we are concerned. Government (1) specifies exactly what kinds of wealth and property shall be money, fixes the monetary unit, defines the different kinds of money and their relations to each other; (2) coins the metallic money and engraves and prints the paper money; (3) in certain cases regulates the quantity of money; and (4) makes and enforces rules regarding legal tender.

Legal tender. While the general acceptability which is the essential characteristic of money grew up naturally through custom,

modern governments have commonly found it desirable to enforce the acceptability of money by law. We thus have *legal tender,* which is defined as *any kind of money which according to law must be accepted when offered in payment of any obligation expressed in terms of the country's monetary unit.*

When, as is usual, the monetary system consists of several kinds of money, the legal tender quality is often not the same with respect to the several kinds of money. The government may bestow legal tender on certain kinds of money and not on other kinds equally serviceable as media of exchange. In the United States prior to 1933, for example, gold coins and United States notes were legal tender, while national bank notes and silver certificates were not. The legal tender distinction is purely a matter of legal enactment and does not necessarily indicate superior intrinsic worth of the kind of money thus designated.

It should be made clear that the legal tender law is not concerned with contracts or other obligations not payable in money. Neither does it apply to contracts which by their terms are payable in a particular kind of money. There is nothing to prevent people making contracts payable in any kind of money they may agree upon. Only when the agreement is in terms of dollars (or other monetary unit), no particular kind of dollars being specified, does the legal tender law apply. It then serves, in case of dispute, to settle just what is meant by the word "dollar."

Since practically all money agreements are thus generally stated or understood, legal tender laws are actually of wide application. Their normal purpose is (1) to protect the creditor from being forced to accept anything but the best kind of money when different kinds are in circulation, and (2) to protect the debtor against a capricious creditor who might refuse to accept a perfectly good kind of money or insist upon receiving some form which the debtor could not conveniently provide.

It may be observed in passing that the government has the power, by altering the monetary unit or amending the law as to legal tender, to introduce unforeseen changes in the fulfillment of contracts, both public and private, previously made in terms of money.

Money and currency. Although, as we have noted, it is the principal function of money to serve as a medium of exchange, the fact is that many exchanges are consummated by means of instruments which are not in the strict sense money. Bank deposits—more

strictly, individual deposits subject to check—are a medium of exchange, and indeed it appears that in the modern industrial nations bank deposits bear the larger part of the burden of this fundamental monetary function.[1] It has been estimated for example that in the United States some 80 or 90 per cent of the total volume of exchanges is conducted by means of checks or similar instruments which transfer bank deposits.

Such instruments do not qualify as money under our definition because they are not "generally acceptable." This makes it convenient to have another term, broader than money, which will include everything that serves as a medium of exchange. The term *currency* has been generally adopted. The complete picture of a *currency system* must include bank deposits and certain other forms of currency, such as postal money orders, express checks, etc. The subject of bank deposits will be treated at length in a later chapter.

In noting the great importance of bank deposits and other forms of currency which are not money, the reader will of course not forget that all forms of currency (except money itself) are themselves rights to receive money; in other words, that the whole currency system is based upon money. This point may be expressed in the statement that, although all currency performs the functions of a medium of exchange and a store of value, only money serves as a standard of value.

Varieties of money and monetary systems. Of the various kinds of money developed from monetary experience, *standard money* emerged first and still holds a place of primary importance in monetary theory. The term is used to designate *those forms of money which are commodities or articles of wealth having value in their own right apart from their use as money. Standard money is money having "intrinsic value" equal to its face value.* In modern times standard money has usually consisted of gold or silver coins or gold or silver bullion.

Standard money is usually accompanied in circulation by certain other kinds of money, whose value is not based upon the materials from which they are made but arises from the right of the holder to redeem such kinds of money in the standard money. To all

[1] This statement is particularly applicable to the United States, Great Britain, and other Anglo-Saxon communities, where the deposit system of transfer by check is very highly developed. In the countries of Continental Europe there is relatively less reliance upon deposits and more upon money.

kinds of money which are freely redeemable in standard money we give the title of *representative money.*

A monetary system which consists of standard money or of both standard and representative money is known as a *standard money system.* This is the system which generally prevailed in most of the civilized world for a century or two preceding the First World War. In the United States a standard money system prevailed, with occasional interruptions, from 1792 until the abandonment of the gold standard in 1933.

In contrast to a standard money system is a money system in which the ultimate monetary unit is not gold, silver, or any commodity of intrinsic value, but a form of money deriving its value not from the material of which it is made but from the fact that it circulates and is generally acceptable in the community. Such a money system is known as an *irredeemable money system.* Historically irredeemable money systems usually developed out of previously existing standard money systems, with the money consisting of government notes or bank notes whose promise to pay was not honored. Many European nations went over to irredeemable money during the First World War, and most of the nations of the world are in one form or another on that basis today.

The quality of general acceptability of irredeemable money may be attributed to several factors. Originally, the paper currency is accepted because of the hope of ultimate redemption in coin or bullion, but regardless of the possibilities of redemption, the paper money continues in use from the sheer force of habit developed before the abandonment of conversion privileges. Finally, and probably of greatest importance, is the sanction bestowed upon the currency by the state by conferring full legal tender qualities upon the paper currency.

Standard money system: Principal kinds of money. In a standard money system there is usually in circulation a greater volume of representative money than of standard money.

Experience has developed several kinds of representative money, and the following classification of the component parts of a typical standard money system may prove useful:

I. Standard money
II. Representative money
 A. Standard money certificates

B. Credit money
 1. Government notes
 2. Bank notes
 3. Token money

A *standard money certificate* testifies to the fact that the government holds standard money which will be delivered to the bearer of the certificate on demand. It is thus of the nature of the warehouse receipt, and the reserve held by the government may never be less than the total amount of the certificates outstanding. The function is solely one of convenience; *i.e.,* to permit the use of the standard money without the necessity of handling the actual coins. The United States gold certificates, prior to 1933, were an example of standard money certificates.

The term *credit money* covers those forms which are of the nature of the promissory note. A *government note* is the government's promise to pay standard money or some other kind of money to the bearer on demand. A *bank note* is a similar promise of a bank. The United States notes (greenbacks) and the federal reserve notes, respectively, are examples in the United States.

Credit money (unlike standard money certificates) does not require the holding by the government of standard money equal in amount to the credit money outstanding; all that is required is a reserve sufficient to meet actual demands for redemption.

Token money is credit money consisting of coins containing less than their nominal value of metal and redeemable in standard money. The token coins are usually made of silver or of other less precious metals, as copper, nickel, or bronze. When made of the same metal as the standard coins, the token coins are of proportionately less weight. Thus there is not as much silver in two half dollars, four quarters, or ten dimes, as in a silver dollar (the latter having formerly been a standard coin). In the case of token money, the promise to pay is not stated upon the coin but is contained in the law requiring the public treasury to redeem the coin on demand of the bearer. Conversely the government will always give out the token coins on request in exchange for standard money.

Token coins are usually of smaller denominations than the monetary unit. Their function is to permit the "making of change"; *i.e.,* payments involving fractional parts of the monetary unit. Ex-

amples are the fractional coins of silver, nickel, and bronze in use in the United States.

The ideal money commodity. Although many different commodities have at various times and places been employed as money, most of these monetary expedients have gradually been eliminated, until in modern times the civilized world has quite generally settled down to the use of the precious metals, gold and silver, as the standard money. Nations have continually sought for their standard moneys a commodity having certain qualities necessary for an efficient medium of exchange, and a study of the evolution of the kinds of standard money reveals these qualities and indicates how well gold and silver meet these specifications.

First of all, the ideal standard money must have *general desirability*. The precious metals have always been desired for the fabrication of ornaments, and in modern times they have also important industrial uses, as in dentistry.

In the second place, gold and silver have *great value in small bulk*. This is an important consideration, since a medium of exchange should be susceptible of convenient use and easy transportation. Consider for example the inconveniences which would accompany the use as money of such otherwise useful commodities as wheat, coal, or salt.

Thirdly, gold and silver have the advantage of *durability*. Age and exposure do not materially affect them, and they may thus be kept indefinitely without danger of loss from physical deterioration. This is of course an important matter, if money is to perform its function as a "store of value." The Virginia colonists for example frequently suffered loss through the deterioration of their stores of tobacco money.

Another desideratum of a convenient money commodity is *uniform quality*. All the possible advantages of monetary exchange are not obtained when each piece of money must be examined and tested as to its quality. Wheat would make a poor kind of money for this reason; so would coal or cotton or indeed almost any other possible commodity. The Virginia legislature had to take cognizance of the distinction between "good tobacco" and "bad tobacco" and provide officials to examine and grade the tobacco. Even so, the colonist had always to be on his guard against receiving payment in inferior or spoiled money. The precious metals, on the other hand, have perfect uniformity. When to this is added

the quality of durability, which we have just noted, it results that an ounce of fine gold is exactly like any other ounce of fine gold, whether it came from a California mine in 1849 or was mined last year in South Africa or the Klondike.

The convenience of a monetary commodity is greatly enhanced by *divisibility*. Thus gold and silver may be divided and made into coins of any desired size without loss of value.

Gold and silver have also the obvious advantage of being easily recognized, of *cognizability*. They are not easily confused with other substances, and counterfeiting is difficult.

Types of standard money systems. There are many forms which a standard money system may take, but for our purpose it will be sufficient to distinguish two main types for detailed discussion. They are (1) the gold standard, including two modifications known as the gold bullion standard and the gold exchange standard, and (2) the bimetallic standard.

The gold standard: Characteristics. Under the gold standard, the standard money consists of gold coins. All other kinds of money which may circulate as part of the system (representative money) derive their value from the fact that they are redeemable in gold. Inasmuch as several modifications have been made in the practices of gold-using countries since the First World War, it is desirable to analyze, first of all, the prewar gold standard before undertaking a study of its present operation.

The characteristics peculiar to the gold standard are: (1) a monetary unit prescribed by law; (2) free coinage of gold; (3) legal tender of gold coins; (4) free circulation of gold, including the right to hoard; (5) unlimited convertibility of all other forms of currency into gold; and (6) the maintenance of a free gold market, which includes the right to import and export gold.

The monetary unit. In a standard money system the law prescribes the monetary unit in which the various kinds of money are stated. The United States was on the gold standard from 1879 to 1933. The unit was the dollar, consisting (after 1837) of 25.8 grains of standard gold, standard gold being a mixture of nine-tenths pure gold and one-tenth an alloy of copper. The dollar thus contained 23.22 grains of pure gold and 2.58 grains of copper. Similarly the British monetary unit before September, 1931, was the sovereign, or pound, containing 123.273 grains of standard gold; British stand-

ard gold being eleven-twelfths fine, the sovereign contained 113.001 grains of pure gold and 10.273 grains of alloy.

Note that the unit in a standard money system is always a certain quantity of metal whose weight and composition are exactly prescribed by law. The unit may or may not be represented by a coin. Very often the monetary unit is too small to make a convenient coin. This is of course the case in the United States.

Coinage. Coinage is the process of making metal into coins to be used as money. The coins bear the stamp or mark of the issuing state.

The government obtains the metal for coinage in one of two ways. Either it buys the metal in the open market, or it allows *free coinage.* Free coinage may be defined as *a system by which the government is legally required to coin for any person any amount of a particular standard metal which he may bring to the mint.* Free coinage is a necessary attribute of standard money. In the United States there was free coinage of gold until 1933.

Free coinage does not mean the absence of a charge for the actual task of coinage. When free coinage is granted without a charge for the cost of the operation, the term *gratuitous coinage* is employed. There has been no charge for coinage under the United States gold standard except during the years 1853 to 1895, when the charge was first ½ of one per cent, and, after 1873, ⅕ of one per cent.

If the gold presented to the mint for coinage is not of standard composition, a charge known as *brassage* may be made to cover the cost of converting the metal so presented into standard bullion.

The value relation between gold and money. Since the standard money under the gold standard consists of gold coins of an established weight and fineness, which the holder is free to hoard or melt down into bullion, the value of the standard money can never fall below that of an equivalent weight of gold in the form of bullion. On the other hand, if the value of gold coin were ever greater than the value of the corresponding weight of gold bullion, dealers, taking advantage of the right of free coinage, would buy gold in the commodity market and have it turned into coin until there was no longer profit in the operation because the values had become the same. Because gold coin and gold bullion can be freely interchanged without loss, the value of standard money tends

always to be equal to the value of the substance of which it is made; *i.e.*, to its intrinsic value.

Under this system the various forms of representative money will have values equal to their nominal values because of the fact that they are all redeemable in terms of standard money, either by the government or by the agency which has issued them.

The gold bullion standard. A modification of the gold standard was employed by France and England for a time after the First World War. For example England, having departed from the gold standard during the war, returned in 1925, not to her former standard, but to what is known as the *gold bullion standard.*

The right of free coinage of gold was given up (except in the case of gold bullion brought to the mint by the Bank of England), and the Bank of England notes and other representative money were denied the right of redemption in gold coin. Instead there was substituted redemption by the Bank of England of its notes in gold bullion at the price of £3 17*s.* 10½*d.* per standard ounce, but only in large bars containing approximately 400 ounces of fine gold. The Bank was also required to purchase gold bullion whenever offered at the price of £3 17*s.* 9*d.* per standard ounce.

We have here illustrated the essentials of the gold bullion standard. There is not free coinage and little or no circulation of standard coins. Representative money is redeemable, not in gold coin, but in gold bullion. Redemption is usually limited to such large quantities of gold that the right is in practice available only for gold export or for very large domestic transactions. The fact remains, however, that there is redemption in the standard metal, and this fact is sufficient to keep all forms of represenative money at par with gold. Free trading in gold bullion is permitted, but the absence of standard coin and the limitation of redemption to such large quantities of money virtually prevents hoarding.

The gold exchange standard. During the period following the First World War, another variation of the gold standard, known as the *gold exchange standard,* was extensively adopted in Europe, especially by small countries. In order to avoid the necessity of holding a large supply of gold as a basis for their currency, these nations made their representative money redeemable, not in gold, but in exchange upon some country having the gold standard. When an individual presented a sum of representative money for redemption, he would receive foreign exchange, which could be

used to obtain gold from abroad. The government itself treated bank deposits in such financial centres as New York, Paris, and London as parts of its own monetary reserve of standard money.

It will be noted that under this standard the currency of a foreign country becomes the standard money. The government is obligated to redeem representative money either in gold or in foreign exchange at its own option. It is usual for the government to keep part of its reserves in the form of gold and part in the form of foreign exchange. There is of course no free coinage of gold, and little or no circulation of gold coin, but redeemability in gold exchange keeps the various forms of representative money at parity with gold.

Although the gold exchange standard has the double advantage of economizing in the use of gold as a reserve and of enabling reserve funds held abroad frequently to earn interest, there are serious difficulties in connection with its widespread use. It is an added obstruction in the already rocky path of the semi-automatic operation of the gold standard. It exposes the monetary and banking system of the depositary countries to the hazard of entirely unpredictable gold drains by the depositor nations. Furthermore it subjects the depositors to the danger of serious losses due to a suspension of gold payments by the country in which reserve balances are deposited.

The bimetallic standard: Description. *Bimetallism,* though no longer in effect in any nation, has played an important role in monetary history and involves economic principles of great interest. In a bimetallic system both gold and silver are concurrently used as materials for the standard money, and the monetary unit is defined as either a certain weight of standard gold or a certain weight of standard silver, the law prescribing also the exact composition of each standard metal. *The ratio between the weight of pure silver in the silver unit and the weight of pure gold in the gold unit is called the mint ratio, coinage ratio, or legal ratio.* There is free coinage of both gold and silver, and the standard coins of either metal are unlimited legal tender.

For example, the monetary system first established by the United States of America (in the year 1792) was a bimetallic system. The standard coins were the silver dollar and certain gold coins. All the coins were unlimited legal tender. The gold dollar contained 24.75 grains of pure gold and 2.25 grains of alloy, a total weight of 27

grains of the standard metal. The silver dollar contained 371.25 grains of pure silver and 44.75 grains of alloy; *i.e.,* 416 grains standard. The coinage ratio was therefore 371.25:24.75, or exactly 15:1. These weights were later slightly changed, and after 1837 the mint ratio was 15.988:1, the later famous "sixteen to one."

Operation of bimetallism. Bimetallism gives rise to certain important questions in addition to those which relate to monometallism. If debts may be paid either in gold coin or in silver coin, and if either kind of coin may be freely obtained at the mint in exchange for the respective bullion, will it make any difference which metal is used, and if so what considerations will determine the choice? Will both metals be actually used?

The theoretical analysis essential to the answers to such questions begins with consideration of the relation between the mint ratio and the *market ratio,* the latter being *the ratio between the value of a given weight of gold and the value of the same weight of silver on the market.* This ratio is of course the result of the respective values of gold and silver, which are determined in their turn by the demand and supply, and demand and supply depend respectively upon the uses of the metals for all purposes (not merely for money) and the conditions of mining. This ratio can not be controlled by law, and it is subject to frequent change, as is the ratio between the values of any two commodities.

At any given time the market ratio must be either equal to, or less than, or greater than the mint ratio. Starting with the first case, let us assume, for purposes of illustration, a bimetallic system with a mint ratio of 16 to 1, and let us further assume that at a certain time the market ratio between silver and gold is also 16 to 1. Under these circumstances, a given amount of gold will exchange on the market for 16 times its weight of silver, and a given amount of gold will make at the mint exactly the same number of dollars in coins as 16 times its weight of silver. In this event there will be no reason for taking one metal to the mint rather than the other, and both metals will be coined.

Taking now the second case, let us suppose that the mint ratio is 16 to 1, while the market ratio is 15 to 1. The market ratio is thus less than the mint ratio, no one will take silver to the mint, and only gold will be coined. Furthermore there will be an advantage in melting down and using as bullion any silver coins which may then be in circulation. Silver coins thus tend to dis-

appear from circulation, gold is the only metal coined, and the only coins in circulation are gold. The effect is practically the same as gold monometallism.

The opposite results follow if the market ratio is greater than the mint ratio; for example, a mint ratio of 16 to 1 and a market ratio of 17 to 1. Only silver will be coined, and the gold coins already in circulation will tend to disappear. The effect is practically the same as silver monometallism.

Gresham's law. It is the metal which the mint overvalues in comparison with the market ratio which drives the other out of circulation. Coins of the first metal may be obtained from the mint more cheaply than coins of the other, for the reason that the mint places a higher value on that metal in terms of the other than does the market.

This principle is part of a broader generalization, known as "Gresham's law," which is fundamental to monetary theory and may be stated as follows: *When two or more kinds of money of unequal value are in concurrent circulation, each being available for payments, the inferior tends to drive the better out of circulation.*

Adjustment of discrepancy between ratios. The forces set in operation by a discrepancy between the mint ratio and the market ratio have themselves a tendency to bring the ratios back to equality. If the market ratio between silver and gold is less than the mint ratio, the existing silver coins gradually pass out of circulation, and new coins are made only of gold. The task of furnishing standard money, formerly shared by gold and silver, thus shifts to gold. This means an increased demand for gold and a consequent tendency for its market value to rise. On the other hand, silver being no longer used for standard money, the demand for silver is less, with a tendency for its market value to fall. The rise in the value of gold and the decline in the value of silver tend of course to increase the market ratio between them and so to correct the original discrepancy.

It is possible theoretically that the equilibrium between the market ratio and the mint ratio might thus be restored before all the standard silver coins had been eliminated. From that point the coinage of silver would be resumed, and the bimetallic system would go on as before, except that some of the silver coins would

be definitely lost and the circulating medium would contain relatively more gold and less silver than previously.

The opposite train of events would occur if the market ratio were greater than the mint ratio; *i.e.,* increased demand for silver and decreased demand for gold, lowering of the market ratio, and theoretical possibility of bringing the two ratios together. Exactly how great a discrepancy between the market and mint ratios could be overcome by the added demand for the overvalued metal depends among other things upon the relative conditions of production of gold and silver bullion and upon whether the burden is borne by a great nation or a small one, by a few nations or by many acting together.

However, the effectiveness of this principle to perpetuate effective bimetallism in the face of a fluctuating market ratio of gold and silver is strictly limited. The United States had a bimetallic monetary system from its first monetary law in 1792 until the year 1873. Never in all those years was the market ratio maintained at equality with the mint ratio. There was in actual effect always either silver monometallism or gold monometallism. France, during most of this period, also had bimetallism, with a mint ratio, however, that was very close to the market ratio. She was for a time successful in keeping both metals in concurrent circulation, but had France not abandoned bimetallism in 1873 she could not possibly have maintained the equality of mint and market ratios in face of the falling price of silver.

We may conclude that the automatic principle will operate practically only at a time when the difference between the mint ratio and the market ratio is extremely small.

Continued interest in bimetallism. Throughout the last quarter of the nineteenth century, after the gold standard had become entrenched in the leading commercial nations, interest in bimetallism continued. The problem was accentuated by the declining price of silver and the rising price of gold at this time. Demand for the restoration of bimetallism arose, coming chiefly from the producers of silver, the debtors who were suffering from the trend of falling prices, and those traders who found that the declining rate of exchange on the silver-standard countries made trading with these countries increasingly difficult.

A common argument advanced in behalf of bimetallism is that it makes a greater quantity of money available and hence checks

a downward price trend. In reply it may be pointed out, first, that there is no scarcity of gold today, only a maldistribution of the stock. In the second place, the solution for monetary crises and falling prices is to restore confidence in the normal credit system, not merely to increase the stock of money.

It has also been claimed that bimetallism would tend to make the value of money more stable than it is under a monometallic system. This argument relies on the assumption that any discrepancy between the ratios would tend to be self-correcting, thus preventing the disappearance of either gold or silver from circulation. This assumption, as we have seen, has not been borne out by actual experience. Moreover it is assumed that a change in the value of one metal would be offset by an inverse change in the value of the other, thus leaving the aggregate value of the monetary unit unchanged. However, there is no certainty that the value of gold and silver will not fluctuate together and thus magnify the currency instability rather than reduce it. Experience has shown that a single country is unable to remain on a bimetallic standard, and bimetallism on an international scale has never been tried.

The last argument in favor of bimetallism that merits serious discussion is that which claims bimetallism would give more stable exchange rates with silver-using countries. Though this argument might have some theoretical validity, it is of no practical consequence today due to the universal abandonment of the silver standard.

Irredeemable money systems: Origin and essential features. The distinguishing characteristic of an irredeemable paper money system is the absence of any standard money whatever. And by definition there can of course be no representative money. The monetary authority, whether it be the government or the central bank, no longer redeems money either in a commodity of intrinsic value or in the money of any other nation.

While it is perhaps conceivable that an irredeemable paper money system might be originally established without reference to a previously existing standard money system, this has as a matter of history never occurred. Actually the origin is to be found in the suspension of specie redemption of the representative elements (generally government notes or bank notes) in a standard money system. This action calls into operation "Gresham's law," the working of which forces out of circulation the standard money and any

other forms of money which may still be redeemable. The irredeemable paper money thereafter becomes the basis of the system.

The world's monetary history is replete with examples of irredeemable money. To mention early experience in America only, irredeemable money was a commonplace among the colonists; the chief monetary medium of the Revolution, the Continental bills of credit, were irredeemable money; and the greenbacks of the Civil War put the nation on an irredeemable money standard from 1861 to 1879.

Up to the time of the depression beginning in 1929, nations had rarely given up their standard money systems except under the pressure of war or of some other crisis. Departure from the standard money system was a symptom of financial weakness. In recent years, however, many nations have abandoned the gold standard in favor of irredeemable money systems, not because they were actually forced to do so or because their metallic reserves were in danger of exhaustion, but for reasons of deliberate policy.

The irredeemable money system, like the standard money system, may comprise several different kinds of money. The basic money will generally be either government notes or bank notes. Frequently both of these forms of credit money will be in use, the bank notes being redeemable in government notes. Token money is generally present, performing the same function as in a standard money system. It is not redeemable in standard money, though it may be redeemable in other forms of money of larger denomination. The token money may be paper notes instead of coin. Bank deposits of course play the same role as in a standard money system, deposits being payable in the irredeemable paper money.

It may at first sight be hard to understand how irredeemable money can have any value at all, or why anyone should consent to receive such money. Some of the reasons were stated earlier in the present chapter, and the fact that people do accept irredeemable paper money is demonstrated over and over in the annals of the world's monetary history. Force of habit and the lack of any better currency continue the use of irredeemable money even when the chance of future redemption has become virtually nil.

Paper money may deteriorate so far in value that it ceases to function as a store of value, yet so long as the paper money still serves its function as a medium of exchange, people will continue to use it. Each individual feels that he can accept the money so

long as he can pass it on to others, and so the one essential characteristic of money, general acceptability, is maintained. The Continental bills of credit of the American Revolution circulated till they were worth less than one cent on the dollar. The common people of Germany after the First World War were still using their paper mark when its value in terms of gold coin was impossible to calculate. When it was finally stabilized, the government gave one gold mark for one trillion marks of the paper money!

General abandonment of the gold standard. Upon the outbreak of the First World War all European countries, neutrals as well as belligerents, abandoned the gold standard in order to safeguard their metallic reserves, and most other countries followed suit. Irredeemable money systems, consisting generally of the notes of central banks, prevailed throughout most of the world.

The United States was one of the few nations still adhering to the gold standard during this period. Even in the United States an embargo on gold exports was enforced from late in 1917 to the summer of 1919, and while redemption of other forms of money in gold was not refused, it was discouraged. During this period therefore the United States was not completely on the gold standard.

The period of recovery between 1923 and 1929 witnessed a concerted return to the gold standard. Certain small countries—for example, Turkey and Spain—failed to take part in this restoration movement, but some thirty other countries did so. England returned to gold in 1925 at the prewar valuation of the pound. Many other nations, such as France, Italy, and Belgium, reduced the gold content of their monetary units upon returning to the gold standard.

The depression overthrew the gold standard again. In 1929 and the years following, one country after another—more than thirty countries by March, 1934—abandoned the gold standard, most of them for the second time. England stood out for many months, but was forced off gold in September, 1931. The United States abandoned the gold standard in 1933. The British dominions, the Scandinavian countries, and certain others subsequently regulated their currencies so as to maintain a determined relationship to the pound sterling and thus came to be known collectively as the sterling area. Only a small group of nations—France, Belgium, Holland, Switzerland, Italy, and Poland—remained in 1934 to consti-

tute the gold bloc,[1] resolved to defend the gold standard against the forces threatening its complete overthrow. However, the maintenance of their old gold parties in the face of about a 40 per cent devaluation of the pound and of the dollar resulted in an overvaluation of the currencies of the gold bloc countries.

The resultant loss of exports, falling prices, and domestic credit deflation finally forced these countries to abandon gold in the autumn of 1936. France left the gold standard and devalued the franc on September 25, 1936, concurrently entering, with Great Britain and the United States, into a tripartite agreement to work together on a day-to-day basis to maintain stable exchange rates for their three currencies. The other gold bloc countries likewise depreciated their currencies, and some joined the other three powers to maintain exchange stability.

Recent American monetary experience: Abandonment of the gold standard. The crisis of 1929 did not develop weaknesses in the American money system as such, but it did accentuate certain defects of long standing in the banking mechanism, which progressed to a state of almost complete collapse late in 1932. Runs upon the banks, withdrawal of money for hoarding by frightened depositors, and wholesale bank failures led the governors of several states to order all their banks closed in February and March of 1933.

On March 5, 1933, only one day after his inauguration, President Roosevelt by executive proclamation ordered all the banks in the United States to be closed on March 6 and took the first steps in abandoning the gold standard. The proclamation of March 5, 1933, went further and forbade, except with special government license, the export of gold or silver or the paying of gold or silver by any bank. Simultaneously the United States Treasury ceased redemption of gold certificates and all gold payments.

Four days later the emergency banking act of March 9 confirmed the President's proclamation and further authorized the President to regulate or prohibit hoarding of gold or silver and to require all persons to surrender to the United States Treasury all gold coin and bullion and gold certificates in exchange for other kinds of money. These provisions as to gold were immediately put into effect. It was generally expected that at the end of the banking

[1] Germany and a number of central and eastern European nations remained nominally on gold but imposed restrictions on foreign exchange.

crisis the nation would return promptly to the gold standard. Acceptance of this understanding is demonstrated by the fact that foreign exchange rates registered no immediate depreciation of the dollar.

Actually, while the acute banking crisis was virtually over in a week, no move was made to restore the gold standard, and on April 19, 1933, the President proclaimed an embargo of indefinite duration on gold exports, having previously extended the regulations against hoarding of gold by individuals. This was formal notice of indefinite abandonment of the gold standard. The proclamation was immediately reflected by depreciation of the dollar on the foreign exchange markets of the world.

The concluding act came on June 5, 1933, when the President approved a joint resolution of Congress which canceled the gold clause in all federal and private obligations, making them, as well as all other contracts and debts, payable in any form of legal tender. This action led to lawsuits ultimately involving some 100 billion dollars in gold-clause obligations. On February 18, 1935, the Supreme Court rendered a 5 to 4 verdict in favor of the government, holding that creditors must accept in depreciated currency, dollar for dollar, the sums named in their contracts, and that, while the government's action in repealing the gold clause in its own contracts was unconstitutional, plaintiff had no case against the government because no actual damage had been shown. The government subsequently enacted legislation designed to guard itself against further suits in this connection.

Devaluation of the dollar. The first Roosevelt Administration was confronted with a host of economic problems. The United States, as well as the rest of the world, was in the throes of a serious depression. The banking system was at the point of collapse. Unemployment was mounting to dangerous heights. Prices, particularly of those staple commodities, such as wheat and cotton, which are sold on an international market, had been declining since the 1929 crisis.

Monetary policy was geared to implement the rest of the Administration's economic program. It was felt in Administration circles that in order to encourage private enterprise and increase employment, prices had to be raised. Therefore it was considered necessary first to cut the monetary system loose from gold and then to depreciate the value of the dollar in terms of gold. Depreciation

of the dollar was sought by means of an extensive program of gold purchases, begun late in October, 1933. The offer by the government to purchase gold with dollars was of course equivalent to an offer to sell dollars for gold, and any buying price set above the former gold standard rate was equivalent to depreciating the gold price of the dollar. The government began to buy newly mined domestic gold at $31.36 an ounce, far above the old gold standard price of $20.67 an ounce. Shortly thereafter this program was extended to include buying gold abroad, and the buying price was successively raised until it reached $34.45 an ounce on January 15, 1934.

The results of the depreciation of the dollar were far from what had been expected. Although by January, 1934, the price of gold was 59 per cent above the pre-devaluation level, wholesale prices had risen only 20 per cent. Even this small rise in prices probably can be attributed to nonmonetary measures of the Administration and to the general upswing in business conditions rather than to the devaluation program.

Gold reserve act and proclamation of 1934: Nationalization of gold. The gold-purchase program was followed on January 30, 1934, by the far-reaching gold reserve act and the accompanying proclamation of January 31, which devalued the dollar and nationalized gold. These measures provide that there shall be no gold coinage, no circulation within the United States of gold or gold certificates, no free redemption of currency in gold, no dealing in or holding of gold except under Treasury license; they empower the Treasury to deal in gold so as to maintain the parity of all kinds of currency, and they provide that gold bullion shall be made available under Treasury license for the settlement of international balances.

The Treasury buys all gold presented, except such as may be known to have been hoarded in violation of the orders established in 1933, subject to the usual mint and handling charges; hoarded gold remains subject to confiscation with a 100 per cent penalty charge, although in practice the Treasury has been purchasing such gold at the former mint price of $20.67.

Title to all gold coin and bullion held by the federal reserve banks was transferred to the Treasury in return for a payment in gold certificates at a rate determined by the old gold content of the dollar. These new gold certificates are held by the banks as final reserves and are redeemable at the Treasury only when gold

bullion is needed to settle foreign balances. The notes of the federal reserve banks are now redeemable in "lawful money," which means other forms of paper money. Coinage of gold has been stopped, and all gold held by the Treasury has been converted into bars.

The new monetary unit. The gold reserve act also prescribed that the weight of gold defined as a dollar should not be more than 60 per cent of the former weight and granted the President continuing power to make successive revaluations within specified limits for a period of three years. This grant of power was subsequently extended until June 30, 1943, when Congress allowed the power to expire.

It was directed that any increase in the value of gold (*i.e.*, the number of gold dollars equivalent to an ounce of gold) then held by the Treasury resulting from a reduction in the weight of the gold dollar should be carried in the Treasury as a miscellaneous receipt and that, in the event of any increase in the weight of the gold dollar, the reserve for notes and the security for gold certificates should be correspondingly augmented by a transfer of gold from the general fund of the Treasury.

The President immediately fixed the gold content of the dollar at $15\frac{5}{21}$ grains, $\frac{9}{10}$ fine, the equivalent of an even $35 an ounce as the dollar price of gold. The gold dollar since 1837 had contained a pure gold weight of 23.22 grains and a standard weight (*i.e.*, $\frac{9}{10}$ gold, $\frac{1}{10}$ alloy) of 25.8 grains. As now fixed, the dollar contains $13\frac{15}{21}$ (approximately 13.7) grains of pure gold and has a standard weight of $15\frac{5}{21}$ (about 15.24) grains. This is a reduction of the dollar to 59.06 per cent of its former weight.

Legal tender. The act of June 5, 1933, which abrogated the gold clause in public and private contracts, enacted that "all coins and currencies of the United States (including federal reserve notes and circulating notes of federal reserve banks and national banking associations) heretofore or hereafter coined or issued, shall be legal tender for all debts, public and private . . ." This makes all forms of money, apparently even the small-change coins, unlimited legal tender.

"Gold profit" and stabilization fund. The dollar value of the Treasury's gold was increased by the devaluation of the dollar to an amount slightly in excess of two and three-quarters billion dollars. The federal reserve banks were left with a claim to as many dol-

lars' worth of gold as before devaluation. But since the new dollar contained only 59 per cent of the former weight of gold, the banks' share of the total gold in the country's monetary stock (measured either by weight or by dollar's worth) was reduced to 59 per cent of what it had been.

The gold reserve act gave the Secretary of the Treasury power for a period, originally of three years but later extended, to deal in gold, foreign exchange and any other necessary instruments, in order to stabilize the exchange value of the dollar. A fund of two billion dollars was provided for this purpose, to be taken from the increased value of the Treasury's gold which might result from devaluation, with the permission that any portion of the fund not being currently used in stabilization operations might be used by the Secretary of the Treasury for the purchase of government obligations.

Silver policy. With the abandonment of the gold standard, the silver issue once more achieved prominence. Agitation for increased recognition of silver in the monetary system was insistent and led to a number of legislative enactments.

The Thomas amendment to the emergency farm relief act in 1933 authorized the President to accept silver up to two hundred million dollars' worth, on the basis of fifty cents per ounce, in payment of the allied war debts due in 1933 and to issue silver certificates against the silver so received. A number of payments received in 1933 were thus made in depreciated silver.

At the ill-starred world monetary and economic conference in London in July, 1933, an agreement was negotiated designed to raise and stabilize the price of silver. As its share of the plan the United States agreed to purchase twenty-four and a half million ounces of silver a year. In effect the agreement called upon this country to protect the price of silver against whatever damage might be done it by the limited disposal of some of India's surplus stock.

In December, 1933, the President, acting under the provisions of the gold reserve act, announced that the United States mints were open, during the next four years, for the purchase of all the silver mined in the country at the statutory price (unchanged since 1837) of approximately $1.29 per ounce, the producer however to receive only half of this amount, or approximately 64.5 cents per ounce, while the balance represented a 50 per cent *seigniorage*

charge, or profit, to the government.[1] This was 21 cents above the current market price at that time.

Demand for further action on behalf of silver persisted and came to a head in June, 1934, when Congress passed the silver purchase act. The Secretary of the Treasury is directed to buy silver as he may deem it advisable at a price not above its monetary value ($1.29), except that in the case of silver in the United States prior to May, 1934, the price may not exceed 50 cents. The Secretary may sell silver whenever its market price is above its monetary value. The purchases are to continue until silver constitutes 25 per cent of the monetary stock of the country. Silver certificates, now made full legal tender, must be issued at a face value not less than the cost of all the silver bought, and they may be issued in an amount not larger than the statutory valuation of the silver.

By the same act the President was given power to combat the hoarding of silver by nationalizing the commodity, which he did by executive order in August, 1934. The order required the delivery of silver bullion to the Treasury within 90 days at 50 cents an ounce. The Treasury promised to see that adequate supplies of silver would always be available at a fair price for legitimate commercial purposes.

So far as the American monetary system is concerned, the action taken with regard to silver has thus far been of little significance. The United States Treasury has acquired a considerable amount of silver bullion. There has been a moderate increase in the circulating medium through issue of silver certificates. However, from June 30, 1933, to April 30, 1948, the stock of silver dollars and bullion in excess of the requirements for backing the certificates increased only 128 millions, and only 1,902 millions, or 4.5 per cent of the total increase of 42,099 millions in the monetary stock of the nation between these two dates, are due to silver acquisitions. Of the silver bullion held by the Treasury on April 30, 1948, there were only 188 million dollars' worth not already pledged to redeem outstanding silver certificates.

The announced one-to-three ratio between silver and gold in the monetary stock is of no immediate practical significance. The present stock of silver dollars and bullion is far short of the amount required and is not likely to reach the announced ratio in the near

[1] A charge made by a government for coining bullion is called *seigniorage* if it exceeds, and *brassage* if it does not exceed, the actual expense of coinage.

future. And in any event the ratio means little in a monetary system which has largely lost connection with the Treasury's stock of gold. Only as presaging a future return to bimetallism is there real monetary significance in the present silver purchase policy. The essential result of the policy is to grant silver producers a government subsidy.

The present American monetary system. The gold reserve act and following executive order completed the process by which the United States definitely abandoned the gold standard and went onto a new money system.

Subject to certain formalities and at the option of the Secretary of the Treasury, gold may be obtained in return for paper money for export and for domestic use commercially and in the arts. The restrictions which are now contained in the monetary system of the United States were designed primarily to make impossible the hoarding of gold, while not interfering unduly with its use for paying foreign balances or in the arts. This involves denial of free coinage and of the use of gold coin in circulation and withdraws the essentials of a standard money system; *i.e.,* free redeemability of all forms of paper money in gold at the pleasure of the holder and a free gold market, both domestic and for export.

The United States Treasury is obliged to purchase gold bullion whenever offered at the rate of $35 an ounce. This of course prevents the value of the paper money from ever rising above that ratio in terms of gold. On the other hand, the right of the citizen to receive gold bullion for export in exchange for paper money at $35 an ounce will keep the money from falling below that ratio in the international market so long as the Secretary of the Treasury, in the exercise of his discretion, permits such redemption without undue restriction. Exchange of paper money for gold for domestic use in the arts, also limited and at the discretion of the Secretary of the Treasury, is clearly not a sufficient redeemability to assure the gold value of the paper money internally, though it does maintain a certain limited relation between the paper money and gold. The present American monetary system, though correctly classified as an irredeemable paper money system, is thus not turned entirely adrift from gold as in the extreme forms of irredeemable paper money systems.

The accompanying table shows the composition of the United States monetary system as it was on June 30, 1947. Comparison with

CIRCULATION STATEMENT OF UNITED STATES MONEY, APRIL 30, 1948[1]

(In dollars)

Kind of money	Held in U. S. treasury	Held outside U. S. Treasury			Total stock
		By federal reserve banks and agents	In circulation		
			Amount	Per capita	
1. Gold	23,168,729,469	——	——	——	23,168,729,469
2. Standard Silver Dollars	336,142,971	2,853,198	154,352,530	1.06	493,348,699
3. Silver Bullion	1,949,125,005	——	——	——	1,949,125,005
4. Token Coins:					
a. Subsidiary Silver	14,800,588	29,923,694	902,844,835	6.18	947,569,117
b. Minor Coins	8,106,415	7,683,586	342,134,275	2.34	357,924,276
Total Coin and Bullion	25,476,904,448	40,460,478	1,399,331,640	9.58	26,916,696,566
5. Gold Certificates	(19,094,123,984)[a]	2,815,444,500	45,615,039	0.31	(21,955,183,523)[a]
6. Silver Certificates	——	245,020,749	2,009,465,166	13.76	(2,254,485,915)[a]
Total Certificates	(19,094,123,984)[a]	3,060,465,249	2,055,080,205	14.07	(24,209,669,438)[a]
7. United States Notes	2,194,292	34,064,856	310,421,868	2.13	346,681,016
8. Treasury Notes of 1890	——	——	1,146,166	.01	(1,146,161)[a]
9. National Bank Notes	361,430	745,750	100,251,391	.69	101,358,571
10. Federal Reserve Bank Notes	442,637	3,764,265	361,311,848	2.47	365,518,750
11. Federal Reserve Notes	50,360,215	907,317,615	23,488,931,905	160.85	24,446,609,735
Total Notes	53,358,574	945,892,486	24,262,063,178	166.15	25,260,168,072
Total Monetary Stock	25,530,263,022	4,046,818,213	27,716,475,023	189.80	52,176,864,638

[a] Not included in the total of this column, in order to avoid double counting.

[1] Source: U. S. Bureau of Public Debt, *Circulation Statement of United States Money, April 30, 1948.*

corresponding figures for June 30, 1932, discloses certain important changes.

The dollar value of the gold holdings of the Treasury is over seven times what it was in 1932 (partly due to reduction of the unit). This is all in the form of bullion. There is no longer any gold coin in circulation. The amount of gold certificates has increased greatly, but these are mostly held by the Treasury and the federal reserve banks and agents; only 46 million dollars are still in circulation. The principal part of the gold certificates in the Treasury is due and payable to the federal reserve banks.

Almost two billion dollars' worth of silver bullion has been acquired by the Treasury. No substantial change has taken place in the silver dollars, but the stock of silver certificates has more than quadrupled. Silver certificates are redeemable in silver dollars, and the Treasury is required to maintain the value of the silver dollars at a parity with other kinds of money, as before. From 1932 to 1948 federal reserve bank notes increased a hundredfold and federal reserve notes eightfold, while United States notes remained at the same amount, and national bank notes were reduced by over 600 million dollars.

The currency system of course includes as always individual bank deposits subject to check. Deposits are payable in lawful money, all forms of money being now legal tender.

EXERCISES

1. The silver dollar contains 371.25 grains of pure silver. There are 480 grains in an ounce.
 (*a*) The market price of silver today is about 75 cents per ounce. What is the intrinsic value of the silver dollar?
 (*b*) At what price of silver would the silver in a dollar actually be worth a dollar?
 (*c*) Suppose silver dollars were recoined into an equal number of dollars of lighter weight, and the surplus silver sold as bullion. How would this affect their intrinsic value? Their value as coins? How would it affect silver producers? Industrial consumers of silver? Explain in each case.
2. The following figures indicate the weight of pure gold in the gold dollar at different periods in our monetary history:

1792–1834	24.75 grains of pure gold
1837–1933	23.22 grains of pure gold
1934–	13.71 grains of pure gold

 (*a*) Determine the mint price per ounce for gold, during each of these periods.

(*b*) Daily quotations are to be found in the newspapers for silver, but no quotations for gold. Why?

(*c*) Prior to 1933, Congress had fixed the weight of the gold dollar at 23.22 grains of pure gold, and the weight of the silver dollar at 371.25 grains of pure silver. The price of gold was fixed, while the price of silver fluctuated. How was this possible?

3. After 1873 the market ratio rose as high as 80 to 1. At that ratio what was the market price of silver? Why did not silver coin drive out gold coin and leave the United States on a silver standard?
4. What would be the weight of the gold dollar if the mint price for gold were fixed at $20 an ounce? $24 an ounce? $30 an ounce? $35 an ounce?

CHAPTER XV

Principles of Banking

Nature of credit: Financing production. To pursue further the process by which goods and services are exchanged under modern monetary systems, we need to inquire into the operations of the banking system. As was stated in the preceding chapter, bank deposits, resulting from credit provided through normal banking processes, constitute in the United States and other Western nations by far the largest portion of the currency. Moreover, in a stage of advanced division of labor, banking provides the machinery for assembling savings and even for creating the credit required to facilitate the production and distribution of the maximum flow of goods and services to satisfy human needs. In these two respects, banking plays a central and critical role in the proper functioning of a free economy.

Let us for the moment assume a very simple type of productive organization. There are only manufacturers, wholesalers, retailers, and consumers. Goods pass in a direct line from the manufacturer to the wholesaler, to the retailer, and to the consumer. There are no banks or other financial houses to assist any one of them. When the manufacturer buys materials and pays wages he must pay in cash. When the wholesaler buys goods from the manufacturer he pays cash for them; the retailer and the consumer do likewise.

The financial functions of these persons seem slight, but are they? The manufacturer is compelled to use his own funds while he is manufacturing the goods; the wholesaler ties up his resources in a stock of goods until he has disposed of them to the retailer; and the retailer is in the same position. Even the consumer may buy somewhat in excess of his immediate needs. As a matter of fact, each one is *financing* himself.

Some of the consumers, let us assume, find it impossible to pay for the goods immediately, and they induce the retailer to "trust" them for small amounts. But as a result of this situation the retailer may find himself unable to pay cash for the goods he buys

from the wholesaler. He is confident that, if he were given three months in which to make payment, he could sell a certain amount of goods and secure payment from his customers, and so he suggests to the wholesaler that he be allowed to purchase the goods on the basis of deferred payments—that he be given three months' credit.

If the wholesaler agrees to this, he assumes the burden of financing the retailer as well as himself. Obviously to do this he must have fairly large resources or be compelled to curtail the extent of his operations, provided he still pays the manufacturer cash on delivery of goods. But he may make to the manufacturer the same plea which the retailer made to him, urging in extenuation his treatment of the retailer.

If the manufacturer accedes to his request, he is financing himself, the wholesaler, the retailer, and the consumer. He is selling goods and receiving in exchange a written or verbal promise to pay for them some time in the future.

Definition of credit. Here we see the essence of the credit system. Credit involves the purchase of something in the present—merchandise, money, or services—agreeing to pay for it in the future. In business the amount and time of the future payment is usually determined, and the payment is almost always made in money or some other form of currency. But this is not necessarily so. A farmer may borrow a few bushels of seed wheat, promising to return a certain number of bushels of wheat when his crop is harvested.

The credit system is conditioned on the willingness and ability of individuals or of business organizations to advance present goods, whether money, merchandise, or services, in return for a promise to pay in the future. It therefore depends on the existence of a surplus, over and above present needs, which can be placed at the disposal of others.

In our present example credit has been entirely of the type we shall call *commercial credit;* that is, it has been extended by the manufacturer, the wholesaler, or the retail merchant. This held true in the main of the industry and commerce of former times.

The financial houses devoted exclusively or primarily to financing businessmen are of fairly recent origin. For although it is true that banks have existed for many centuries—the Bank of Venice was founded in the twelfth century—and that we find money lenders and financial houses of importance in the medieval period, they

were the exceptions to the general rule, and the producers were forced to rely on their own or their families' resources or on each other. It was not until the eighteenth and nineteenth centuries that banks and other institutions so increased in number and resources as to become a vital part of the economic fabric and that *bank credit* (a promise by a bank to pay) surpassed in importance what we have here called commercial credit (a promise by a manufacturer, wholesaler, or retailer to pay).

The function of credit: Credit in general. Although bank credit has become of predominant importance in the economic organization, it would be an error to ignore the service of commercial credit; *i.e.,* credit extended by the manufacturer to the wholesaler, or by the wholesaler to the retailer. In many transactions the banker is not called upon for credit. The manufacturer is able to take care of himself, or the producer from whom he gets his raw materials gives him time in which to work them up and pay for them.

Credit, whether commercial or banking, is a sort of solvent factor in our organization, enabling us in a sense to make use of future goods at the present time. For society as a whole this cannot hold true of course, but it can for a given individual, for he gets present goods in return for a promise to pay in the future, and the transaction is conditioned on the belief of the lender in the borrower's ability to produce goods or services at some time in the future.

Credit, which enables some to make present use of the surpluses of others, facilitates that smooth working of the productive apparatus which is essential to economy. It is one of the factors which promote efficiency in production and which therefore tend to keep the output of consumable goods at a high level.

Bank credit: Pooling of savings. Bank credit performs this function in society better and more cheaply than other kinds of credit. The resources of any chain of producers are great, no doubt, but they are limited when compared with the resources of society as a whole, including the active leaders of business, those who are active in business as wage earners, and those who live on incomes derived from investments. The banks stand at the central point of the financial world and serve as reservoirs which attract the savings of all. They are in touch with more people and more lines of industry than the typical firm can possibly be. If there is for the time being

a surplus of funds in one line of industry, it can be made available for other industries.

Without the bank as an intermediary, how for example would the proprietor of a cotton mill in Massachusetts know that the owners of brass factories in Connecticut had funds which would be idle for six months? And if he did discover this fact, how could such funds be made available to him? It might be done, but it would be difficult and costly.

The bank stands ready to receive the funds of the Connecticut manufacturer and to pay him interest on a time deposit for six months; it can loan these directly to the cotton manufacturer in return for a promissory note of six months' duration. The thing is accomplished directly, simply, and at little expense. The accomplishment of this result is clearly an aid to production and so a service to society as a whole, not merely to the lenders and borrowers directly concerned.

In performing the function of gathering together the savings of the community and making them available for use in production, the banks effect an important economy in the use of the community's savings. If every producer and consumer had no recourse to borrowing from the banks and had to keep on hand the funds to meet his greatest needs, much of the savings of the community would necessarily be hoarded and withheld from use in production. It is through the pooling of savings by banks, thus counterbalancing the high and low demand for savings of the different individuals in the community, that the production system is enabled to obtain a much increased use of such savings as are available.

Creation of credit. But while the banking organization discharges an essential economic function as an assembler and distributor of savings, it is also the creator or manufacturer of credit, and in this role its activities are, as we shall see, both more significant and quantitatively more important. Through its processes of making loans and receiving deposits, the bank takes the credit of an individual, with its limited acceptability, and transforms it into a form of credit with much wider acceptability, which in Anglo-Saxon countries particularly has attained a position of predominance in the currency system. In the remainder of this chapter, we shall inquire into the principles and activities by which the banks perform these functions.

Types of banking institutions. Many types of banking institu-

tions have developed during the history of banking, to perform in whole or in part the work of assembling and distributing savings and of creating credit. Some of these institutions, such as the mutual savings banks, receive, hold or invest, and pay out upon application the savings of their customers. Trust companies are for the most part also commercial banks, but their characteristic function is that of handling estates and otherwise acting as trustees and executors. Note brokers, commercial paper houses, and finance companies are essentially middlemen participating, through buying and selling credit instruments of one type or another, in the process of providing funds to business firms.

It is the commercial banks however which, through the processes of what are called discount and deposit, play the predominant role in modern economic organization. These banks specialize in loaning to business firms funds which they have obtained through the deposit with them of savings or through their own creation of credit, and these activities give rise to the demand deposits which figure so significantly in the currency system. So important are the functions of discount and deposit that the commercial banks performing these functions are appropriately distinguished from all other banking institutions by being designated banks of discount and deposit.

Loans and discounts: Credit creation through lending. How bank credit comes into existence through the lending operations of a commercial bank may best be shown by following through in detail a loan transaction of a hypothetical moderate-sized city bank. Let us assume that J. C. Miller, a dry goods merchant in the town, is engaged in buying his stock of goods for the early spring trade. As is usual, he does not have enough capital of his own for the purchase of his entire stock but depends regularly upon the financial assistance of his bank. He therefore goes to the bank and asks to borrow $50,000. The officers of the bank, after satisfying themselves as to his character and the state of his business, agree to

$50,000	New Haven, Conn., October 28, 1947

Sixty days after date I promise to pay to the order of the City Bank fifty thousand dollars. Value received.

(Signed) *J. C. MILLER*

accept his promissory note for $50,000 payable, let us say, sixty days from date. This note would be in some such form as the note shown at the bottom of page 284.

Interest and discount. The compensation which the bank gets for supplying Mr. Miller with funds is called interest or discount, and the process is called loaning or discounting, depending upon the way the note is drawn.

If the note is drawn *with interest;* that is, if Mr. Miller promises to repay at maturity the $50,000 plus interest at, let us say, 6 per cent for 60 days, he is said to be getting a *loan* of $50,000 and to be paying *interest* on $50,000. In this case, at the time of borrowing he receives $50,000 and at maturity will pay $50,500. The bank properly regards the promissory note as worth $50,000 at the time of borrowing and so enters it among its assets.

If however Mr. Miller writes his note for $50,000 *without interest,* the note is not worth $50,000 this day to the bank, and it must determine what value to allow him for it. The bank knows that 60 days from date it will bring $50,000; according to the principles of valuation with which the reader is now familiar,[1] it might be supposed that the bank would determine what sum put at interest for 60 days at 6 per cent would amount to $50,000; namely, $50,000 ÷ 1.01, which is $49,504.95. The $495.05 represents the *true discount,* or the amount by which the face value exceeds the present value of the note.

Banks are accustomed however, because of the extra profit and the easier computation, simply to deduct interest at the agreed rate on the face value, leaving in this case $49,500. In banking parlance the $500 is called *bank discount* or simply *discount,* and the $49,500 is called the *proceeds.* Mr. Miller in this case really procures only $49,500 from the bank but pays interest on $50,000. According to custom, the bank enters this note among its assets at $50,000, which involves an immediate overvaluation of $500. This is corrected by adding $500 to an item, "unearned interest," on the liability side of the bank's statement. There is transfer from this item to undivided profits as notes approach maturity.

The term *discount* is used both to refer to the interest deduction and to describe the whole process of evaluating and disposing of the note. In the financing of commerce the bank discount method

[1] See Chapter XIII.

is most common at the banks, but there are many types of financing in which the loan method prevails.[1]

Proceeds. The next question is as to the form in which Mr. Miller shall receive the funds borrowed. There are ordinarily three forms: (1) money (other than the bank's own notes), (2) the notes of the bank, (3) deposits. The first is simple enough and needs no explanation. In the second case the bank gives Mr. Miller, in exchange for his promissory note, its own promissory notes. It is a swapping of notes. But the notes of the bank are money; they are generally acceptable throughout the community, and Mr. Miller has received what he wanted, the means for making immediate payment for his goods. Indeed the ordinary business man does not distinguish between two different kinds of money; to him there is practically no distinction between the first two cases. To the bank, as will appear later, there is a very real distinction. Both of these forms of payment however are the exceptional cases, at least in the United States. In the great majority of instances the borrower receives the proceeds of his note, not in money ("cash"), but in the form of a deposit credit.

When money is taken to a bank and left—"deposited"—the bank credits the customer's account on its books, and the latter obtains the right to withdraw the specified amount of money, or any part of it, at will. Now when a note is discounted the amount of the proceeds is (in the ordinary case) credited to the account of the borrower as a deposit, precisely as though he had brought money. If it is any aid to visualizing the process, it may be imagined that the customer received the proceeds in money and immediately stepped over to the receiving teller's window and deposited the money. Actually however no money has been involved; this is indeed not a money transaction. It is once more a swapping of credit. The bank receives the customer's promissory note and gives him in exchange the right to receive money from it on demand, which is essentially what he would have if he had taken instead the bank's notes.

Principal types of trade paper. We must now retrace our steps in order to consider the several forms of paper which may be used in ordinary commercial borrowing at a bank. There is first the

[1] As is the case with most economic categories, the terms "loan" and "discount," "interest" and "discount," are loosely and interchangeably used by the public.

simple promissory note of the borrower. Such a note may be *unsecured,* or it may be *secured* by "collateral security." In the latter case the borrower delivers to the bank along with the note certain property of his, worth at least as much as the amount of the note, with an agreement that if the note is not paid when due, the bank may sell the security and so reimburse itself for the note. Stocks and bonds of corporations and government bonds are the property most used as collateral security, though warehouse receipts for certain commodities, bills of lading, etc., may be used.

In the example which we have used, Mr. Miller discounted at the bank a simple promissory note, either secured or unsecured, and obtained the $49,500 which he used to pay the wholesaler. The purchase of his stock in trade might, however, have been somewhat differently financed.

It is common practice for the wholesalers to sell to their customers on credit; that is, to accept their promissory notes in place of cash. If Mr. Miller has such an arrangement with a wholesaler, R. B. Norton, in the same city, he will give Mr. Norton in exchange for the goods his promissory note for $50,000 payable in 60 days. The wholesaler, Norton, being himself actively engaged in business with all his capital, will not be willing to hold Miller's note till maturity; he needs the cash before that and will therefore avail himself of the services of the bank. Norton first *endorses* Miller's note; that is, signs his own name, thereby making himself liable for its payment in case Miller should not pay. He then takes the note to the bank, which discounts it precisely as when Miller borrowed on his own note. The proceeds are given to Norton, either in money or, more likely, as a credit to his deposit account.

The note becomes the property of the bank, and when it is due Miller will pay, not Norton, to whom he gave the note originally, but the bank. A note thus endorsed is called "two-name paper" as compared with the simple promissory note of the maker, called "one-name paper."

The paper offered to the bank may take still a third form; namely, the *draft* or *acceptance.* The retailer, instead of giving the wholesaler his promissory note in return for goods received, may give him the authority to *draw* upon him. This, in the case we are using for illustration, would mean that when Norton sent the goods to Miller he would write out an order upon Miller directing him to pay the $50,000 in sixty days to Norton's bank. This draft would

then be taken to Miller, who would *accept* it; that is, write the word "accepted" and his signature upon its face, which is his acknowledgement of his obligation to pay at the maturity of the draft. This draft would be in form similar to that here shown:

> New Haven, Conn., October 28, 1947
>
> To J. C. Miller
>
> On December 28, 1947, pay to the order of the City Bank fifty thousand dollars ($50,000).
>
> The transaction which gives rise to this instrument is the purchase of goods by the acceptor from the drawer.
>
> Accepted at New Haven, Conn.
> October 28, 1947
>
> Payable at the City Bank,
> New Haven, Conn.
>
> (signed) *J. C. MILLER* (signed) *R. B. NORTON*

A draft when thus accepted is called an *acceptance.* It is two-name paper, since the drawer, Norton, is liable as well as Miller, upon whom it is drawn. Mr. Norton then takes the acceptance to the bank and discounts it precisely as has been explained in the case of Miller's promissory note discounted by Norton. At the maturity of the draft, Miller will pay the bank. These various forms of businessmen's notes and drafts are all called *trade paper.*

A bank's balance sheet. The essential nature of this loan, as well as lending in general and other typical banking operations, may be more clearly comprehended by tracing through at this point the development of our bank's balance sheet and, in the process, considering the effect of Miller's loan on its financial condition.

Since a commercial bank is ordinarily a corporation, its first balance sheet upon organization might read as follows:

STATEMENT OF THE CITY BANK, OCTOBER 1, 1947

Assets		*Liabilities*	
Cash	$400,000	Capital stock	$400,000

The bank proceeds at once to buy land and a banking building, furniture and supplies, etc., for $60,000, and thereafter its balance sheet reads as follows:

STATEMENT OF THE CITY BANK, OCTOBER 15, 1947

Assets		*Liabilities*	
Real estate	$ 45,000	Capital stock	$400,000
Furniture, etc.	15,000		
Cash	340,000		
	$400,000		$400,000

Discounting a note. Now comes Mr. Miller, the merchant, with his request for a loan. As has been already assumed, he offers his promissory note for $50,000 payable in sixty days, the bank's rate of discount is 6 per cent, the discount is $500, and the proceeds, $49,500. Mr. Miller elects to receive "cash" other than the bank's own notes.

The effect of this transaction upon the bank's condition is shown in the following statement:

STATEMENT OF THE CITY BANK, OCTOBER 28, 1947

Assets		*Liabilities*	
Loans and discounts	$ 50,000	Capital stock	$400,000
Real estate	45,000	Unearned interest	500
Furniture, etc.	15,000		
Cash	290,500		
	$400,500		$400,500

It will be noted that the $50,000 promissory note acquired by the bank is listed among its assets opposite the heading "loans and discounts," that the cash has been reduced by the payment of $49,500 to Mr. Miller, and that the difference between these amounts, the discount, appears as unearned interest. Obviously the effect on the bank's statement will be the same whether Mr. Miller borrows by means of an unsecured note or a note secured by collateral. Even if Miller finances his business by giving his promissory note to the wholesaler or permitting the wholesaler to draw upon him, the final result is the same, the only difference being that in these cases (that is, the endorsed note or the acceptance) Miller's obligation is discounted at the bank by Norton, the wholesaler, instead of by Miller himself. In any case the bank comes into possession of Miller's obligation, which he must pay to the bank at maturity.

The statement will, however, differ according to the form in which the proceeds are given to the customer. The statement above

illustrates the case where the proceeds are taken in the form of money other than the bank's own notes. If the bank had given its notes, the statement would have read thus:

STATEMENT OF THE CITY BANK, OCTOBER 28, 1947

Assets		*Liabilities*	
Loans and discounts	$ 50,000	Capital stock	$400,000
Real estate	45,000	Unearned interest	500
Furniture, etc.	15,000	Notes	49,500
Cash	340,000		
	$450,000		$450,000

If finally the proceeds had been credited to the deposit account of the borrower, the statement would have been as follows:

STATEMENT OF THE CITY BANK, OCTOBER 28, 1947

Assets		*Liabilities*	
Loans and discounts	$ 50,000	Capital stock	$400,000
Real estate	45,000	Unearned interest	500
Furniture, etc.	15,000	Deposits	49,500
Cash	340,000		
	$450,000		$450,000

Paying a loan. Now let us suppose that December 28 has come around; this is the date on which Mr. Miller's loan matures and must be paid at the bank. This is equally the case whether Mr. Miller borrowed directly from the bank on his own promissory note or gave his promissory note to the wholesaler, Norton, or accepted a draft drawn upon him by Norton, so long as the paper has been discounted at this bank.

Mr. Miller now presents himself at the bank and offers to pay his note. He may pay in any of five ways: (1) in money, other than the notes of this bank, (2) in notes of this bank, (3) by a check upon a deposit account in this bank, (4) by a check upon an account in some other bank, or (5) by a bank draft, *i.e.*, a check drawn by a bank either on itself or on its account in some other bank. The effect upon the bank's condition is readily determined. In any case the loans and discounts will be reduced by $50,000, representing the note paid and returned to Mr. Miller. If the note is paid in money, the bank's cash, or reserve, will rise by $50,000, and the same result will follow, within a few days at most, if payment is

made by a check or a bank draft on another bank, since this will be promptly presented to the other bank, through the clearing house or by mail or messenger, and collected. If the note is paid by giving notes of this bank, the note item on the liability side of the account will be reduced by $50,000. Finally if payment is by a check either on Mr. Miller's account or on the account of another depositor in this bank, the bank's deposits will be correspondingly reduced. Two changes of $50,000 each counteract each other, and the statement remains in balance.

The interest of $500, which the bank had anticipated but had not yet earned when it discounted Miller's note on October 28, becomes actual earnings of the bank when the note is paid on December 28. The bank accordingly transfers on its balance sheet this amount from the unearned interest account to an account designated "undivided profits."

A typical bank statement. To carry our illustration somewhat further, let us assume that during the next few months the bank has done a flourishing business, making loans, discounting paper, and correspondingly increasing its deposits and notes and profits, that a part of the latter has been transferred to surplus, and that the bank has made some investments in securities. The statement, corresponding more closely to a normal banking business, may be assumed to be as follows:

STATEMENT OF THE CITY BANK, JUNE 30, 1948

Assets		*Liabilities*	
Loans and discounts	$1,250,000	Capital stock	$ 400,000
Securities	450,000	Surplus	75,000
Real estate	45,000	Undivided profits	8,117
Furniture, etc.	15,000	Unearned interest	4,210
Reserve	182,460	Notes	50,000
		Deposits	1,405,133
	$1,942,460		$1,942,460

The reader will note the appearance in this statement of a new term, the *reserve*. This is an inclusive term indicating the sum of such assets as are readily available for payment of the bank's liabilities, especially its notes and deposits. The term is usually synonymous with cash or items equivalent to cash. American banks do not generally use this term, specifying rather the separate items which compose their legal reserves.

Bank deposits: Definition. The second principal economic function of the commercial bank is that of deposit. *A bank deposit is a right to receive money from a bank, evidenced by an entry on the bank's books and by the customer's passbook, duplicate deposit slip, or other device.* Deposits are classified as demand or time deposits, the bank being obligated to pay the former on demand of the depositor but obligated to pay the latter only after a certain number of days' notice. Deposits may result from the discount of trade paper as illustrated above, or from deposit of money, or from deposit of checks drawn by another depositor, either of this or another bank, or in certain other ways that need not concern us here.

To illustrate the deposit of money, let us now suppose that a customer brings in $10,000 of money for deposit. The bank's statement immediately thereafter will read thus:

STATEMENT OF THE CITY BANK, JUNE 30, 1948

Assets		*Liabilities*	
Loans and discounts	$1,250,000	Capital stock	$ 400,000
Securities	450,000	Surplus	75,000
Real estate	45,000	Undivided profits	8,117
Furniture, etc.	15,000	Unearned interest	4,210
Reserve	192,460	Notes	50,000
		Deposits	1,415,133
	$1,952,460		$1,952,460

It will be noted that nothing on the bank's statement distinguishes between deposits obtained by paying money to the bank and deposits obtained by the process of discount. There is no difference. What the depositor has is exactly the same in either case, not money, but the right to receive money.

Deposits are not, as is sometimes popularly supposed, "money in the bank." Banks never have on hand a stock of money equal to their deposits. That could hardly be expected since, as we have seen, deposits may originate without any money being given to the bank. Even when money is deposited, the money ceases from that moment to be the property of the depositor; it belongs to the bank and may be spent by the bank as it sees fit. The depositor has simply the right to receive a certain amount of money from the bank.

Discount and deposit. In fact both discount and deposit, contrary to the meanings popularly attached to the terms "loan" and

"deposit," are actually exchanges, transfers of certain rights for other rights or for money. When the bank discounted Mr. Miller's promissory note, it acquired a property right against Mr. Miller and in exchange gave him a property right against the bank; namely, a deposit. Quite properly the bank lists the note as an asset and the deposit as a liability. On the other hand, Mr. Miller enters the note on his accounts as a liability and the deposit at the bank as an asset. When Mr. Norton, having accepted Mr. Miller's promissory note in payment for goods, discounted the note, he sold it to the bank in exchange for a deposit, or perhaps for some form of money. Money deposited at the bank is, so to speak, "spent"—paid for a deposit.

The double function of discount and deposit is ordinarily a swapping of credit—the person's promissory note or draft or acceptance for the bank's credit in the form of notes or deposits. If it be asked wherein lies the advantage of merely exchanging one form of credit for another, the answer is that what the bank's customer wants is a medium for making payments. His own promissory note will not be generally accepted; the bank's note will. He has exchanged a form of credit generally unknown and unacceptable for the well-known and acceptable credit of the bank. The same is true of deposits, since practically all the customer's payments may readily be made by check. Herein is the fundamental characteristic of bank credit and of the service rendered by the commercial bank.

The check: Definition. The right to receive money which a bank deposit gives the depositor is exercised by means of the check. If Mr. Miller desires to withdraw fifty dollars from the bank in order to make a certain payment or for pocket money, he makes out an order upon the bank, *i.e.*, a check, in the following form:

51–100

No. *731*

New Haven, Conn., *June 30,* 1948

THE CITY BANK

Pay to the order of.....*Myself*.....*$50.00*

.....*Fifty*.....Dollars.

(signed) *J. C. MILLER*

He presents this order to the bank, which thereupon gives him fifty dollars in money and deducts fifty dollars from the amount due him on its books. It will also deduct fifty dollars from its reserve item in the assets column.

If Mr. Miller wished to pay a doctor's bill of fifty dollars, he might obtain from the bank fifty dollars by means of a check like the above and then take the money to the physician. But a simpler method would be to order the bank to make the payment direct to the physician. For this purpose a check would be used in this form:

51–100

No. *732*

New Haven, Conn., *June 30,* 1948

THE CITY BANK

Pay to the order of.....*Robert Brown*.....*$50.00*

.....*Fifty*.....Dollars.

(signed) *J. C. MILLER*

From these examples the nature of the check is apparent. *A check is a written order by a depositor directing his bank to pay money to some person.* The person to whom payment is to be made may be the depositor himself or the bank or any specified third party or the bearer.

Payment by check. Within the bank. Let us follow the history of this check. The check is handed or sent by mail to Dr. Brown, who takes it to the bank and demands the money.

Certain essential facts must be established before the paying teller of the bank will hand over the money. First, he must be satisfied as to the genuineness of the check, which is usually sufficiently attested by the signature of the maker, with which the teller is familiar. Next, he must be certain that Mr. Miller actually has as much as fifty dollars to his credit on the bank's books. The teller must also be certain that it is Dr. Brown himself who is asking for payment; if he is not personally known he will have to be identified. Before payment is made the check must be endorsed; that is, the payee, Robert Brown in this case, must sign his name on the back of the check. Endorsement is the payee's acknowledg-

ment that he has received the money; it makes the check evidence of payment for the bank and for the maker.

Having made the payment, the bank debits the amount to Mr. Miller's account, and eventually, when the account is balanced, the canceled check will be returned to him to be kept as evidence of the payment. The effect of this operation will be to alter two items on the bank's balance sheet, making it read, as compared with the last statement (on page 292), as follows:

STATEMENT OF THE CITY BANK, JULY 2, 1948

Assets		*Liabilities*	
Loans and discounts	$1,250,000	Capital stock	$ 400,000
Securities	450,000	Surplus	75,000
Real estate	45,000	Undivided profits	8,117
Furniture, etc.	15,000	Unearned interest	4,210
Reserve	192,410	Notes	50,000
		Deposits	1,415,083
	$1,952,410		$1,952,410

Instead of demanding the money, the payee may "deposit" the check. He endorses the check and takes it or sends it by mail or messenger to the bank, with the request that the amount stated be credited to his deposit account. The bank must be certain as to all the facts except identification and must be satisfied as to the genuineness of the endorsement. It will then credit the account of the payee (*e.g.,* Robert Brown) and debit the account of the maker (*e.g.,* J. C. Miller). The result of this transaction is exactly as though the payee had collected the money from the bank and then immediately deposited it. The balance sheet of the bank is not changed at all.

The check thus furnishes a means of making payments so convenient, safe, and generally advantageous that, in the Anglo-Saxon countries at least, it is availed of very generally, not only by businessmen but by others. Deposit accounts are kept at the bank. Money when received is regularly deposited at the bank, and the same is true of the far greater volume of receipts which come in the form of checks. Conversely fully four-fifths of all payments are made by means of checks rather than with money. The bank thus becomes an agent for making payments between the people by the transfer of deposits from one to another.

Collection between banks: In general. We have proceeded thus far upon the assumption of one bank with which all persons con-

cerned have their dealings. As a matter of fact in every place of any size there will generally be found two or more banks. Moreover the useful device of payment by check has not been permitted to remain a merely local convenience but has absorbed the great mass of payments between persons in different cities, states, and even nations. This introduces certain complications, which must now be investigated.

When one receives a check, the obvious thing is to take it to one's own bank and either deposit it or ask for the money. This is done regardless of whether the check happens to be drawn on this same bank or upon another. In the first case, the matter is simply arranged, as has been explained. Even in the second case, the bank, though under no obligation to cash or credit the check, will ordinarily do so as an accommodation to its customer, sometimes making a small charge for performing this service. Having given its customer the cash or credited his account, the bank thereupon becomes possessed of the check with its claim upon the other bank. If the other bank is in the same place, the check will ordinarily be presented and paid within twenty-four hours. A longer time may be required when the debtor bank is in another city. The other bank of course then debits the account of its depositor who drew the check.

Thus the check system for payment of money or transfer of deposits operates when there are two or more banks in the town. Daily each bank sends a messenger to each of the others and collects for all checks of each other bank which it has paid.

But in a large city with many banks this business of collecting by messenger becomes troublesome and expensive. In a city with fifteen banks there might have to be 15×14, or 210, separate calls each day. Sooner or later the point is reached where it is worthwhile to avoid this great labor and expense by organizing a clearing house.

Clearing house. A clearing house is an association of the banks of a city through which all the claims and obligations between the banks for paying each other's checks and drafts are settled. While the fundamental nature of the clearing house and its business is everywhere the same, there are numerous variations of detail. The following account is based primarily upon the methods employed in the New York Clearing House.

The business of clearing takes place daily at certain hours.

Before each clearing there occur in each bank the same preparatory operations. The clerks gather together all checks of the other banks which have been received since the last clearing and arrange them in bundles, one for each bank. A list is made showing the amount thus due from each of the other banks, with the total due from all the banks. Sometimes there is also made out a ticket for each of the other banks showing the amount due from that particular bank.

Then two clerks, gathering together all the bundles, the tickets, and the list, proceed to the headquarters of the clearing house. Here there is a room containing a raised desk at one end for the governor and his assistants, and on the floor a row or rows of desks, one for each member bank. Of the two clerks one takes his place behind his own bank's desk while the other, carrying the bundles of checks and the tickets, stands in front of the desk. The list showing the amount due this bank from all others has been sent up to the governor's desk. In like manner, each of the other banks is represented by two clerks, one behind and one in front of its own desk.

At the exact moment for beginning the operation of clearing, a signal on the governor's desk is sounded. The clerks in front then walk along the row of desks, depositing at each the bundle of checks and the ticket against that particular bank. In a minute or two the checks and tickets have thus all been distributed, and it takes only a few minutes more for each desk clerk to add up the tickets or the sums on the outside of the envelopes and so ascertain the amount his bank owes to all of the other banks. These results are reported to the governor, whose assistants are able quickly to discover whether any error has been made. As soon as the figures, corrected if necessary, are compiled, the governor announces the result. Each bank now knows exactly how much it owes all the other banks and how much all the other banks owe it.

These two sums are regarded respectively as a debt to the clearing house and a credit due from the clearing house. The balance represents an amount which the bank must pay to the clearing house or receive from the clearing house, as the case may be. Thus by one payment either to or from the clearing house, each bank settles its accounts with all the other banks. Of course the amount payable by all the debtor banks to the clearing house is exactly equal to the amount payable by the clearing house to all the creditor banks, and the clearing house itself comes out even each day.

Clearing for outside banks. All the banks of a city do not necessarily belong to the clearing house. An outside bank will either make its collections by messenger and make its payments to messengers sent by the other banks in the primitive way, or it may arrange to clear through another bank which belongs to the clearing house. In this case the agent bank presents as its own the claims of the outside bank and accepts as obligations against itself claims of the other clearing house banks against the outside bank.

The volume of clearings. The volume of the business thus simply and smoothly effected by a large city clearing house is enormous. In the New York City Clearing House there were exchanged, during the month of December, 1946, checks and similar obligations to the amount of $34,497,251,355. For the year 1946, the New York City clearings were about $366,000,000,000. The New York City Clearing House stands far ahead of any other in volume of clearings. For the whole of the United States the clearings during the month of December, 1946, were about $66,000,000,000, and for the whole year 1946, $696,000,000,000.

Since our available sources of such statistics do not include reports from all the cities having clearing houses, the figures here given for the whole country are materially less than the actual totals of all clearings. One can imagine what a gigantic task would be imposed upon the banks were they still compelled to rely upon the primitive method of exchanging their obligations.

The reader should be reminded that even these figures do not give a true picture of the total volume of banking business. In fact the recent movement toward bank consolidation, while accompanied by increased business, has obviously tended to a diminished use of the services of the clearing houses. A more reliable indication of banking business is presented by the figures of debit check transactions, which for the year 1947 were 400 billion dollars for New York City and 999 billions for the principal cities of the country.

Out-of-town checks. Checks are regularly used for payments in other towns. When goods are purchased by mail order, payment is ordinarily made by a check drawn by the purchaser upon his own bank and sent by mail to the seller, who may be a thousand miles away. The maker of the check gives himself no further concern over the matter. The receiver of the check deposits it in his own bank, receives credit for it, and concerns himself no further.

His bank has the task of collecting from the other distant bank. Thousands of payments are thus made every day, and every day the mails carry great numbers of checks back and forth all over the country. There is thus imposed upon the banks a service of great importance and enormous volume. An elaborate system has been developed whereby this business is handled with swiftness and certainty and the minimum of expense and, by offsetting checks against each other, with remarkably little actual transfer of money.

The portfolio. A commercial bank as a privately owned, profit-making business seeks to put its assets at all times to the most profitable use. In doing so, it not only advances funds to business as loans and discounts but also invests any funds not otherwise used in stocks, bonds, and mortgages, and in income-producing investments which are readily convertible into cash without loss. To this entire collection of assets, consisting of (1) loans and discounts, (2) investments, and (3) readily marketable securities, is given the designation of the bank's *portfolio.*

All the operations of the commercial bank bear in one way or another on the management of its portfolio, since through the making of various kinds of investments the bank earns its profit. To obtain the largest profit possible but to avoid at the same time an inadequacy of cash to pay deposits on demand, leading to failure, the bank must so manage its portfolio as to achieve (1) liquidity, (2) solvency, and (3) earnings.

However, these three aims are not altogether reconcilable. Liquidity, which is required for short-run solvency, means holding low-yielding, readily-marketable securities, which produce small earnings. Long-run or ultimate solvency can be assured without liquidity by holding, for example, well-secured real estate mortgages which, even if certain of ultimate payment, may not be readily sold before maturity without a loss. Finally earnings, which under a private enterprise system must be sufficient to keep the bank in operation, cannot be sought at too great a risk to liquidity and solvency.

Increasing importance of investments. Although securities, principally bonds, have always been considered desirable bank assets because of their ready marketability, they have become since 1929 a particularly important source of bank income. A combination of circumstances attending the depression resulted in burdening the banks with cash in amounts far in excess of banking requirements.

The sharp decline in business activity reduced greatly the de-

mand for bank loans. The uncertain business prospect discouraged activity on the security markets, and loans made to finance the buying and selling of securities declined so greatly, both in actual amounts and relative importance, that they constitute even now but about 2 per cent of all banking assets. Moreover, an increasing number of business firms turned to financing themselves with funds held as reserves or obtained through the sale of securities and thus became less dependent upon banks for short-term loans.

At the same time, the banks themselves were deterred by business uncertainties from lending their own funds and gradually became dependent primarily on securities, particularly government bonds, as a means of employing their excess reserves. Such investments, which in 1914 constituted only 27 per cent of total loans and investments of all active banks in the United States, and only 30 per cent in 1929, had risen to 78 per cent in 1944.

Since the end of the Second World War, however, bank investments have tended to revert to loans, but in the form of relatively long-term, rather than short-term loans. A large part of the steady and substantial increase in bank loans which has occurred since May, 1945, has been long-term loans. Large corporations have secured loans from commercial banks with terms running up to ten years, in amounts as large as 50 million dollars, and with interest rates as low as 1½ per cent. If this tendency to extend the period of commercial loans continues, it will introduce substantial changes both in the functions of the commercial banks and in the relation of their activities to the monetary system.

The reserve: Its function. The deposits and notes of a bank represent claims against the bank held by outsiders and payable in cash on demand. The bank must be ready to pay any note holder or depositor any amount to which he is entitled at any time. Refusal to pay means "suspension" or failure of the bank. For this purpose the bank must have always on hand a sufficient stock of cash, appropriately called the reserve. But no commercial bank ever holds a reserve equal to the total of its liabilities. In the last illustrative statement (page 295), the bank's liabilities payable on demand (notes and deposits) appear as $1,465,083, for which there is a reserve of $192,410—a *reserve ratio* of a little more than 13 per cent.

In fact the reserve will seldom be as much as 25 per cent and may be very much less. Evidently a simultaneous demand from all

depositors and noteholders would wreck the bank. But of course no such simultaneous desire for payment is to be expected, and the bankers learn by experience about what demands may be anticipated from day to day. They make provision for a corresponding reserve with a proper margin for emergencies.

Since the establishment of the Federal Reserve System in 1914, minimum reserve ratios have been established by law for member banks. Additional reserves are held in the form of till money and reserves to support bankers' balances in inter-bank deposits. These reserves are frequently called "primary reserves."

All banks are from time to time subject to irregular and seasonal drains, and they must be prepared to increase their cash holdings quickly, easily, and without loss. This they can do by including among their assets loans and investments of a highly liquid nature in sufficient quantity to provide against all normal cash requirements that might possibly arise. The only real requisite is that they be easily marketable without loss. Today short-term Treasury obligations serve as a prime source of bank liquidity. Such assets are referred to as "secondary reserves," and vary in amount with the needs and experience of each individual bank.

Control of reserve ratio: Profits vs. safety. The banker is ultimately influenced in his actions by two conflicting motives—profits and safety. To earn profit he must invest the bank's funds, but whether he makes loans or buys securities he not only assumes the risk of nonpayment but also reduces his cash with which alone he can pay depositors on demand. But as loans and discounts are made, cash is immediately withdrawn, or deposits are created, or the bank's note issues are increased. Thus the reserve ratio falls, and the bank's condition becomes proportionately less safe. To the banker's desire for profits is thus opposed the necessity of keeping a safe ratio between the reserve and the demand liabilities.

Under changing conditions the relative strength of these two motives varies. During the 1930's, for example, banks had such large excess reserves that they could greatly expand their loans without endangering their reserve ratios, but during the Second World War these excess reserves disappeared almost completely by reason of the tremendous expansion of loans to business and the government.

Moreover, the shift from loans to investments among the earning assets of the banks has undoubtedly lessened the importance

of the ratio of cash to demand liabilities as an exclusive index of short-run solvency. Yet aside from the emphasis given to the reserve ratio by legislative restrictions and government regulation, it still remains true that the bankers must so dispose of the bank's funds in seeking profits as to assure the continuing ability of the bank to meet in cash its obligations when they become due and payable. For short-run solvency, the reserve ratio is an all-important banking consideration.

Means of control. There are five principal means of controlling the reserve ratio: (1) by the purchase and sale of securities, (2) by rediscounting commercial paper, (3) by borrowing from another bank, (4) by calling in loans which are payable on demand, and (5) by making it more or less easy or advantageous to borrow. The last may be accomplished by changing the rate of discount, by simple refusal to loan, by moral suasion inducing applicants not to apply for loans, by raising or lowering the collateral security requirements, or by requiring more or fewer names on the paper.

Buying and selling securities. As we have seen, a bank usually invests a part of its resources in corporate securities, the main purpose being to derive some income upon funds which if kept as idle cash in the reserve would be earning nothing, while at the same time having something which may be converted into cash whenever needed. The bank therefore chooses for investment securities which may be sold at short notice.

In the statement on page 295 the bank has a reserve ratio of a little more than 13 per cent. Let us assume that this is considered too small and that securities carried on the books at $100,000 are sold for that amount, reducing the stocks and bonds to $350,000 and increasing the reserve to $292,410. The reserve ratio is now $\frac{292,410}{1,465,083}$, or almost exactly 20 per cent. The bank has sacrificed the future interest or dividends on $100,000 of securities, but it has made its position materially safer. Conversely a purchase of securities would reduce the reserve and consequently the reserve ratio.

Rediscounting and borrowing. Instead of selling securities, the bank may strengthen its reserve ratio by selling some of the paper in its portfolio. This is called *rediscounting,* which means to discount again. Some investor is found, either another bank or a broker, willing to take certain promissory notes or drafts at a price

determined by discounting them at a certain rate for the time that must still elapse before their maturity. Suppose our bank selects from its portfolio a note for $60,000, originally acquired two months ago and payable three months from its date, having therefore still one month to run. If another bank is ready to rediscount it at 8 per cent, the City Bank will receive for it $59,600; that is, the face value less one month's interest at 8 per cent. The result of this operation, starting from the statement on page 295, will be reflected in a new statement, as follows:

STATEMENT OF THE CITY BANK, DECEMBER 30, 1947

Assets		*Liabilities*	
Loans and discounts	$1,190,000	Capital stock	$ 400,000
Securities	450,000	Surplus	75,000
Real estate	45,000	Undivided profits	8,117
Furniture, etc.	15,000	Unearned interest	3,810
Reserve	252,010	Notes	50,000
		Deposits	1,415,083
	$1,952,010		$1,952,010

It will be noted that loans and discounts have been decreased by $60,000, the reserve is increased by $59,600, while the discount is reflected as a loss, decreasing the unearned interest by $400. The reserve ratio is now $\frac{252,010}{1,465,083}$, or something more than 17 per cent, instead of 13 per cent as it was before, showing that the desired result has been accomplished.

Similarly a bank may be able to obtain cash by borrowing from another bank on its secured or unsecured short-term promissory note. Referring again to the statement on page 295, we see that if $50,000 in cash were secured in this way the reserve would be increased to $242,410, and the reserve ratio would be $\frac{242,410}{1,465,083}$, or something over 16 per cent, as compared with 13 per cent.

Calling loans. In contrast to the loans which have a definite date of maturity, there are loans which are made payable on demand. If a bank chooses to call for the immediate payment of such loans, it can increase its reserve ratio. If payment is made in cash or in checks on another bank the reserve will increase, while payment in notes of this bank or in checks on this bank will decrease notes and deposits; in either case the reserve ratio rises.

Control by the rate of discount. Another method of regulation is by varying the rate of discount. As has been observed, in the course of a bank's daily business new loans are constantly being made and notes held are constantly maturing and being paid. There is thus an inflow and an outflow of paper, which under ordinary circumstances may about offset each other, keeping the amount of loans and discounts fairly constant though the individual notes held are always changing. Now the payment of loans on hand is already determined by their respective maturities, but anything that can be done to speed up or slow down the rate of making new loans will have power sooner or later to change the amount of loans on hand. For accomplishing this purpose one instrument which the banker has always at hand is the discount rate.

Reverting once more to the statement on page 295, let us suppose that the banker decides that his reserve ratio, 13 per cent, is too small. He has been discounting, let us say, at 5 per cent, and for some time the new discounts have about offset those that were maturing. He therefore raises his discount rate to 6 per cent.

Since he is now charging a higher price for the service of discounting, fewer people will avail themselves of his services in accordance with the law of demand, and the making of new discounts will fall off. The old loans and discounts continuing to mature as before, the net result will be a decline in the bank's loans and discounts. Furthermore, note that when customers pay off their notes they must pay either (1) in cash or in checks on other banks which will promptly be converted into cash, or (2) in the notes of this bank, or (3) in checks on their deposit accounts in this bank. Since loans and discounts are being paid faster than new ones are being made, the bank thus finds its reserve increasing and its demand liabilities decreasing. The reserve ratio thus rises.

Conversely a reduction of the bank's discount rate may stimulate business and cause new loans and discounts to come in relatively faster than the old are being paid off.

Two qualifications should be noted in this connection. In the first place, while the bank has positive power to restrict loan expansion, its power to promote expansion is passive; it can make the borrowing terms more inviting to applicant borrowers, but the decision whether to borrow or not rests with the applicant. Oftentimes, especially in the later stages of business depression, low dis-

count rates and other favoring acts of the banks do not avail to start revival. This principle was illustrated by the generally unsuccessful efforts to stimulate business by easy money and easy bank credit during the years from 1929 to 1939.

In the second place, control of the discount rate is not, in real fact, an important method in the case of regular commercial and industrial borrowers at the banks. The banks use other methods of discouraging or encouraging applicants. When money is tight, commercial and industrial borrowers are able to get funds, if at all, at practically the same rates as when money is easy. Interest rates, except in highly organized markets, are very lethargic. The central bank rate and the rate for certain types of paper broadly dealt in do vary, with the purpose of control, but ordinarily and in the ordinary bank the rate of discount tends to be customary and stable.

The normal business of banking. Acquaintance has now been made with the more important operations which make up the normal daily business of the ordinary commercial bank. The bank's portfolio of loans and discounts contains numerous notes and acceptances of individuals and corporations, to which new notes are being added daily and from which notes are being as constantly withdrawn when they mature and are paid. The bank's deposits are being constantly added to as new loans are made or as customers bring in for deposit either money or checks upon other banks, while, on the other hand, withdrawals by check are continually tending to reduce deposits, particularly in connection with the payment of loans. The bank's notes outstanding tend likewise to rise and fall in harmony with the loans and discounts. From time to time securities held by the bank are sold or new ones purchased, with corresponding changes in the "securities" item.

In connection with many of these operations changes are continually occurring in the bank's reserve, as cash is either received or paid out. So long as the business is profitable, the "undivided profits" item gradually increases, reduced at intervals by the declaration of dividends or by transfer to surplus.

The reader cannot have failed to observe the close relation between the loans and discounts and the notes and deposits. Ordinarily the customers' borrowing increases all these items. Even if cash is taken instead of a deposit credit, the cash tends to come back and increase deposits as it is deposited by those persons to whom the original borrower made payments. When loans are repaid, there

is normally a decrease in loans and discounts and in notes and deposits.

These relationships are fairly obvious in the case of a bank which is alone in its community but the same results will follow for a banking system as a whole, with several banks in many different communities, though there may be considerable shifting of cash and deposits as between different banks and different regions.

Note issue: Similarity to deposits. The reader who has any practical knowledge of banking will not have failed to observe that in this discussion we are concerned primarily with the development of the fundamental principles of banking rather than with the description of any particular banking system. This distinction appears with particular sharpness in connection with the subject of note issue. This banking function has been so much regulated by legislation in the different countries, particularly in the United States, that the reader may have been surprised at the somewhat casual way in which we have thus far treated bank notes, as though they were about the same thing as deposits. As a matter of fact, the essential identity of bank notes and deposits resides in the fundamental nature of banking; it is only legislation that has sharply distinguished them. This point needs some further elaboration.

The bank deposit has already been defined as a right to receive money from a bank, evidenced by an entry on the bank's books and by other means. *A bank note is a bank's promise to pay money on demand.* The term "money" in both these definitions is of course exclusive of the bank's own notes. It will be noted that these definitions are almost identical. Deposits and notes are both forms of bank credit; both are liabilities of the bank; both are (except in the special case of time deposits) payable in money on demand.

When a bank makes a loan, it is immaterial to it whether the credit given the borrower is in the form of notes or deposits. To the borrower it is merely a question whether he shall receive the evidence of his right to receive money from the bank as an entry on the bank's ledger and in his own passbook or inscribed on little pieces of paper which he will take away with him. These differences are superficial, not fundamental.

Differences between notes and deposits. There are however real differences between notes and deposits. The right to receive money from a bank, if in the form of a deposit, may be transferred only

by check. Now a check is of itself no obligation of the bank. The person who receives it must be sure (1) that the signature upon the check is genuine and (2) that the maker actually has at least that amount on deposit and that it will still be on deposit when he presents the check to the bank for payment. Otherwise the check is not good.

Checks are thus available for use only among people who are acquainted with one another or at least have confidence in one another. Their availability is also limited to those who are in the habit of doing business with banks. Checks are thus not generally convenient for payments between strangers, for paying traveling expenses, or for making ordinary wage payments. Where they are available, checks are usually a safer and more convenient means of payment than money. But after all they are not money, since they are not generally acceptable.

The bank note, on the other hand, is money. It is the definite obligation of a bank and, barring the risk of counterfeit or of the failure of the bank, may be accepted without any question as to the honesty of the one who tenders it. The bank note is generally acceptable; it may be used for all kinds of payments.

Bank notes again are expressed in round sums, convenient for payments of all amounts. Although representing rights to receive money from the bank, bank notes may pass about for weeks or months or years before being presented to the bank for payment. They may travel far away from the bank of issue.

The check, on the contrary, usually appears for the purpose of a particular payment; it is expressed in an odd amount of dollars and cents; and its life is ordinarily limited to that one transaction, after which it returns to the bank and is cancelled. That is, deposits require generally a new check for each separate transfer.

These distinctions between notes and deposits, though not so fundamental as their similarities, are yet real and important and justify the issue of bank credit in the two forms.

Legal restrictions. An important corollary of the differences between notes and deposits is that the opportunity for abuse and fraud on the part of an unsound or dishonest bank is far greater in the case of notes than of deposits. This is in part the explanation of the prevailing attitude of government toward the banking functions of deposit and note issue. In spite of their fundamental similarity and their original status as common law rights, the legislation

of most countries has drawn a sharp distinction between them. While the banks have generally been permitted to handle their deposits with little interference by the government, the function of note issue has almost everywhere been subject to strict regulation.

In most countries the right of note issue is restricted to certain kinds of banks or even to one particular bank. Thus in the United States (since August 1, 1935) only the federal reserve banks issue notes. In most of the leading nations of Europe, such as Great Britain, France, Germany, etc., the business of note issue is virtually the monopoly of a single great central bank, more or less closely associated with the government.

Legislation further puts a limit to the amount of notes which the banks may issue, prescribes certain forms of security and a certain amount of cash reserve, determines the form and denomination of notes, etc. As a result the issue of bank notes is in most countries a narrowly restricted and somewhat artificial function as compared with the business of deposit.

The problem of elasticity: Need of elastic currency. A peculiar and important function of bank credit is to furnish an element of elasticity to the currency system. No community's need of money is uniform at all times and seasons. In an agricultural community, for example, the volume of exchanges is especially great in the late summer and fall, on account of the business of harvesting and marketing the crops. Unusually large sums are being paid in wages to farm laborers, the farmers' receipts are at a maximum, and the local merchants find their sales swelling, while at the same time their credit customers are paying their bills.

In a manufacturing town the weekly pay envelopes present a special need of money on Saturdays. The payment of salaries on the first of the month and the prevailing habit of settling accounts at the stores and elsewhere during the first few days of each month make a monthly variation in the need of currency. Corporations very commonly make their payments of dividends and interest quarterly, usually on the first of January, April, July, and October.

There are also the cyclical fluctuations, the successive periods of business activity and business depression, of "good times and bad times." Perfectly to meet this need, the volume of currency should automatically increase and decrease in relation to these normal changes in business activity, so that no changes in the general price level will result.

Elasticity of bank credit. Recalling the classification of money set forth in the foregoing chapter, and the essential characteristics of the several kinds of money, the reader will realize that no item in the standard money system, except bank notes, is so related to the state of business as automatically to expand and contract with the fluctuating need of currency.

With bank credit (notes and deposits) the situation is different. The demand for loans at the bank arises directly out of business transactions, either recently completed or contemplated for the immediate future. When business is active the banks are called upon for many and large loans. And as the banks respond to the demand for loans they increase their issue of notes and their deposits. Currency thus expands to meet business needs. On the other hand, when business declines, the customers of the banks have less need for currency. They pay off their obligations at the banks faster than new loans and discounts are made. In paying they either return to the banks the notes previously issued or they draw checks which reduce deposits. Currency contracts in response to declining need. Evidently there is here that close and automatic connection between volume of currency and business activity which is the mark of an elastic medium of exchange.

It is desirable that elasticity characterize both bank notes and deposits. Many people do not have bank accounts or do not understand the use of checks; in certain regions banks are few and scattered and checks are awkward to use, and for small retail purchases bank notes are more handy. Inasmuch as the volume of all such transactions fluctuates, the needs of business would be served by a system of bank notes that varied with the needs of the people. Actually, in the effort of certain governments to make bank notes safe, their natural elasticity has been destroyed.

The elasticity of deposit credit is limited by the necessity which is upon the banks to maintain adequate reserves. As has been seen, expansion of the banks' notes and deposits reduces their reserve ratios and inclines them to slow down the rate of expansion by raising the rate of discount. On the other hand, contraction of demand liabilities raises the reserve ratio, and the bankers seek to stimulate business by lowering the rate of discount. Banking reserves thus act, not to destroy the normal elasticity of bank credit, but to put on the brakes and so to prevent extremes of either expansion or contraction.

The oversimplified examples given earlier in this chapter have shown that a deposit of cash in a bank enables the bank to increase its loans and therefore its deposits by something—depending on the reserve it is required to maintain—in excess of the amount of cash deposited. Thus if the legal reserve is fixed at 10 per cent of deposits, a deposit of $1,000 in cash will theoretically enable the bank of deposit to support total deposits of $10,000 (the $1,000 deposited plus $9,000 in additional deposits).

While this might be true for the moment, it could not long remain true. The depositor may be presumed to intend to use his deposit to write checks in favor of others having deposit accounts in other banks as well as in this bank. Eventually through the process of transfer by check this deposit of cash will become spread throughout the banking system, permitting each bank that receives additional cash to increase its own deposits. It is therefore correct to say that for the banking system as a whole the deposit of additional cash permits an expansion of loans and consequently of deposits by whatever multiple the legal reserve ratio permits.

If however the volume of bank credit were rigidly determined by the money available for reserves to the individual banks, the currency system would be seriously handicapped in meeting the needs of business. In a general way bank credit does maintain a reasonably constant relationship to the volume of money in circulation. However when business is slow and inactive, cash reserves accumulate in the banks, while in periods of active business the need for cash for payrolls and other purposes expands both loans and deposits. Were the banks entirely dependent upon the fortuitous variations in their own cash, the heavy withdrawal of cash in periods of high business activity might seriously restrict their capacity to create the credit required by business.

It is because the individual banks under such circumstances can replenish their reserves by borrowing from the central bank that this difficulty is overcome for the individual bank. However for the banking system as a whole, there is an ultimate limit to the expansion of bank credit, for central banks also have to maintain reserves. Yet sufficient elasticity in the expansibility of bank credit does exist in most banking systems to meet the legitimate needs of business.

Banking influence on economic fluctuations. Ultimately the banking system, through the assembling and distributing of savings

and the creating and transforming of bank credit into currency, exerts great influence upon the stability and prosperity of the economy, In a later chapter it will be shown that there exist relationships among the variations in the volume of currency, the general level of prices, and the volume of trade. Since stability of the price level is an aid to maximum production and trade, the banking system, through its power over the volume of currency, exerts a significant influence on general business conditions.

For example, failure to provide the necessary credit in a period of expansion could, in the short run, prematurely curtail business activity to the extent of causing a recession, while over the long run it might retard economic progress and the natural development of a nation. These conditions are avoided by an elastic system of bank credit.

On the other hand, experience has shown that during the upswing of the cycle, the optimistic attitude that prevails tends to cause borrowers and bankers to expand the volume of bank credit at a rate which leads to inflation and over-investment and hence hastens the crash. Again, the pessimism of a depression era may result in such a shrinkage in the volume of bank credit that prices and business activity are unduly repressed. Here is illustration of certain tendencies inherent in a banking system which is so elastic that it easily responds to the cyclical fluctuation in business activity.

The power which the banks have to expand their notes and deposits to an amount five to ten times greater than the funds that are deposited with them may have tremendous repercussions on the entire economic order. When a bank gives a borrower bank notes or a deposit, it increases the quantity of purchasing power in circulation and thus enables him to secure present goods and services on the basis of his promise to pay. The newly created purchasing power may affect prices, production, and the distribution of income. The owners of the new funds may use them partly to purchase present goods and services and hence to bid up the price of existing supplies. The increased currency in the hands of those who receive the newly spent funds augments their demand for goods, and thus production is stimulated. The rise in prices, if long continuing, may mean a decline in the real incomes of those persons whose money incomes are constant, which is another way of saying that a change takes place in the distribution of income.

Formerly banks acted as a significant force in directing the flow

of funds. If funds were needed more urgently in one industry than in another, they were directed by the banks toward that industry. The greater the need, the higher the rate of interest paid to secure a loan, and other things being equal, a bank in making its loan gave preference to those willing to pay the higher rate of interest. But with the decline in recent years in the importance of commercial loans and the tendency of firms to finance themselves, this function of the banks has decreased in relative significance. Banks are now engaged principally in the financing of the government deficit, and they exercise no control over the way in which these borrowings are spent.

EXERCISES

1. Draw up a bank statement, using the following items: stock and bonds, $80,000; undivided profits, $30,000; reserve, $50,000; notes, $75,000; loans and discounts, $350,000; real estate, etc., $100,000; surplus, $50,000; capital stock, $100,000; deposits, $325,000; other assets, $15,200; unearned interest, $15,200.
2. Assume that a bank discounts a note for $5,000 for three months, the bank's rate of discount being 6 per cent. Payment is taken in cash. What changes would this transaction make in the bank's statement?
3. Suppose that the proceeds in Exercise 2 were taken as a credit on the customer's deposit account. What changes would the transaction make on the bank's statement?
4. *(a)* What is the reserve ratio of the bank in Exercise 1?
 (b) What is the bank's reserve ratio after the transaction of Exercise 2?
 (c) What is the bank's reserve ratio after the transaction of Exercise 3?
5. The bank of Exercise 1, after discounting the note as in Exercise 2, holds it one month and then rediscounts it with the federal reserve bank, the rate of the latter bank being 5 per cent. The proceeds are taken in notes of the federal reserve bank. Draw up a new statement showing the bank's condition after the operation. What is now the bank's reserve ratio?
6. The clearing house association in a certain city includes five banks. On a given day they present checks at the clearing house as follows:

Checks presented by:	Drawn against:				
	Bank A	Bank B	Bank C	Bank D	Bank E
Bank A	——	$2,100	$1,800	$2,400	$1,600
B	$1,900	——	1,500	2,100	1,800
C	2,300	1,600	——	1,800	1,500
D	2,600	1,800	2,000	——	1,700
E	1,400	2,000	1,400	1,900	——

Determine the amount to be paid or received by each bank.

CHAPTER XVI

The American Banking System

Introduction. The banking system of the United States is the result of continuous experimentation by private individuals and the governments of colonies, states, and the nation, in furnishing money and credit to the people; it bears the earmarks of struggle and compromise among the powerful forces that have featured American national development. Chief among these forces have been insistence upon democratic ideals and local self-government, with corresponding reluctance to concede authority to the federal government, and the fear lest control of the institutions of money and credit be monopolized in the hands of a few. Lessons in sound banking principles have been learned slowly and only after painful and disastrous experience.

The general result emerging from this background is a loose decentralized system of thousands of small local unit banks, each being chartered and supervised by one of the forty-eight states or the federal government.

Historical résumé: Colonial banking. For the beginning of state-chartered banking, we must go back to colonial times. The so-called "banks" of the American colonial period were not real banks, but merely "batches of paper money," as one caustic historian puts it; they did not engage in the business of discount and deposit, and even the issue of their notes did not conform to the principles of banking which our study has developed. But during the Revolution and the years immediately following, a few real banks were founded, which in addition to issuing notes made a start in the field of discount and deposit.

The first Bank of the United States. In 1791, two years after the formation of the national government, Congress established a central bank, the first Bank of the United States. This bank performed the normal functions of a central bank, was the fiscal agent of the United States Government, and did a regular banking business with the general public in competition with the growing company

of local banks. Its notes became the principal circulating medium of the country. It was wisely and conservatively managed, performed its various functions well, and exercised also a salutary influence over the local banks, whose tendency to wild and unsound methods was apparent from the first. Indeed it was this check upon the ambitions of the local bankers and its competition with them that finally proved the bank's undoing. Opposition arose, and Congress eventually failed to renew its twenty-year charter, which expired by limitation in 1811.

For the next five years the nation's banking was wholly in the hands of local banks, chartered by the several states. These banks increased prodigiously in number. Relieved of the restraining influence of the great central bank, the majority ran wild in the paths of reckless and unsound banking. Finally in the financial panic of 1814, the whole banking structure went to pieces, a condition made doubly disastrous as it came in the midst of the War of 1812.

The second Bank of the United States. The people learned the lesson; their misgivings against a strong central bank were temporarily overcome; and in 1816 the second Bank of the United States was founded. This bank's record is in its essential features a repetition of that of the first Bank of the United States, including its notable services to the government and the community and its final engulfment in the bog of politics. Its charter was not renewed, and it closed its doors in 1836. There was now a repetition of the experiences of 1811–16, and the banking system collapsed once more in the panic of 1837.

Decentralized local banking. Re-establishment of a central bank proving politically impossible, the United States Government cut itself loose from banking altogether by the establishment, first in 1841 and finally in 1846, of the Independent Treasury for the custody of all government funds and the management of the government's financial business.

The public was left to rely for banking services upon the local banks. Gradually some order came out of the chaos, and in the older and more settled parts of the country, particularly New England, New York, and Indiana, many sound banks arose and became more or less closely associated in a sound and efficient banking system. Wildcat banking continued to prevail in the frontier regions, and in these parts the people suffered grievously, especially

from the multiplicity of note issues of all degrees of goodness and badness. Thus things continued down to the time of the Civil War.

Legislation during the Civil War introduced a new class of local banks, the national banks, receiving their charters from the federal government. By an act of 1865 Congress imposed an annual tax of 10 per cent upon the notes issued by state banks, making note issue unprofitable, and thus depriving the state banks of that banking function, which thereafter became the exclusive prerogative of the national banks. With this exception, the national banks were an addition to the local banking structure rather than a substitution for the state banks. The latter have continued down to the present day, comprising on the average about two-thirds of all local banking institutions, state and national.

The state institutions consist of state banks, trust companies, and savings banks. The method of organization, the minimum amount of capital required to start a state bank, the nature and scope of banking operations permitted, the reserve ratio and the form in which reserves must be maintained, and the kind of legal supervision over banking and similar matters are laid down in the banking laws of each state, or in the regulations issued in conformity with the laws by the respective state banking commission. In addition there are certain unincorporated private banks.

The national banking system. The Civil War ushered in a new chapter in American banking history. After the first few years the organization of the new national banks proceeded rapidly, and they soon assumed, alongside the state banking institutions, a position of importance in the American banking system, which they have since maintained.

Note issue. The purpose of Congress in inaugurating the national banking system was primarily to aid the financing of the war and only secondarily to reform the banking system. To enlist the aid of the new national banks in the sale of bonds issued to finance the Civil War, the national banking act (as amended after the Civil War) gave to the national banks the exclusive right to issue bank notes and further required that each of the new national banks before commencing business, purchase bonds of the United States and deposit them with the Treasurer of the United States, the amount depending on the capitalization of the particular bank. Upon the security of the United States bonds so deposited, the

bank was permitted to issue its notes in any amount not in excess of the par value [1] of the bonds or the capital of the bank. National banks under certain rare circumstances were also authorized to secure their notes by the deposit of "lawful money."

It will be observed that the notes of the national banks were not issued freely upon the basis of their general assets, in accordance with the simple banking principle discussed in the preceding chapter, but were secured by a particular class of assets set aside for the purpose with a trustee.

The law further provided for the safety of the national bank notes by imposing upon the United States Treasury the obligation to redeem them on demand, of course with the Treasury's consequent right to collect the amount from the bank of issue. For this purpose each bank was required to maintain a cash fund with the United States Treasury equal to 5 per cent of its outstanding notes.

While the device of bond security made the national bank notes eminently safe, this was at the expense of elasticity. Since government bonds had to be bought and deposited before additional notes might be issued, the banks could not quickly expand their note issues in time of need. This was especially true in later years; for as the national banking system expanded and the national debt was reduced, all the bonds with the circulation privilege appearing on the market were snapped up by the banks, and their prices were driven above par. On the other hand, the artificially valued bonds could not be sold in large volume to purchasers except other national banks save at a serious discount, and the banks were for the most part unwilling to hold such low-rate bonds without having the full amount of notes outstanding. One can readily see that there was no likelihood of large contraction or expansion of the volume of these notes in relation to business activity.

The reader will not fail to recall, however, that during the last quarter of the nineteenth century the nation was passing from the use of note currency to deposit currency, and that the importance of an elastic system of bank notes was gradually decreasing.

Government depositories. Besides providing for exclusive note issue by the national banks, the law also provided for relaxation of the Independent Treasury system to the extent of permitting de-

[1] Originally and up to 1900 the limit was 90 per cent of the par value or the market value, whichever was smaller.

posit of government funds in selected national banks and the receipt and disbursement of national bank notes by the Treasury.

Reserves. The national banking act required each national bank to maintain a legal reserve against deposits, the ratio of reserve to deposits depending upon the population of the place where the particular bank was located. However for all banks except those in certain large cities, a portion of this reserve might be in the form of deposits with national banks in specified larger cities.

These arrangements for national bank reserves failed to provide the elasticity of currency so urgently needed in the United States, with its great extent of territory, diversity of industries, and predominance of agriculture. The United States has been *par excellence* the country of small independent banks; by 1913 the nation's banking business was in the hands of some 26,000 local banking institutions, with an average capital of about $80,000 and average resources of about $775,000. Branch banking was until recently forbidden by law to the national banks and in most states to the state banks and trust companies. In the absence of any system of centralized banking reserves, such as has been provided in most countries by the central bank or by a small group of large and powerful banks with their far-reaching system of branches, each bank was forced to rely chiefly upon its own resources to meet whatever situation might confront it.

Struggling to cope with the situation as best they might, the banks (both state and national) in the country districts and smaller towns and cities developed the custom of depositing as much of their resources as they felt they could spare in periods of slack business with banks in the large cities of their respective regions. These city banks in their turn made deposits in other larger cities.

Thus were set up, as it were, little rivulets of currency springing from sources all over the country, uniting in the cities to form larger streams, joining again into still larger streams, till eventually a mighty flow of money poured into the banks of a few financial centers, the greatest of all being New York City. This flow of deposits was encouraged by the willingness of the large city banks to pay interest at a moderate rate upon deposits of other banks. The funds so obtained were loaned by the city banks to their customers, a considerable part being used for stock market speculation.

On the other hand, when a succeeding period of business activity opened up larger opportunities for the outlying banks, or when a

crisis threatened, they demanded the return of their reserves, and the flow of currency was reversed. This forced the city banks to call in their loans and contract their deposits and caused a state of monetary stringency in these cities. This ebb and flow, besides its minor variations, had a great annual cycle because of the need of currency for harvesting and marketing the crops.

It will be seen that, in a sense, the great city banks, particularly in New York, undertook to perform the services of a central reserve reservoir. But they were fatally handicapped by their great number, their independence of each other, their lack of any branch organization, and their inability to expand and contract their credit in the form of notes. Moreover, they were unwilling to accept the responsibility of their position to the extent of keeping adequate reserves.

The result was inevitable. Alternating periods of surplus currency and monetary stringency were the normal thing; inability of the outlying banks to obtain the return of their reserves in time to avoid disaster was common; bank failures and business failures were numerous; financial panics and crises were made more frequent and more severe. The situation grew steadily more intolerable with the development of the country's wealth and industry, until at last a remedy was found in the Federal Reserve System, which was established by law on December 23, 1913.

The Federal Reserve System: Organization: Board of Governors. The Federal Reserve System was established to furnish the United States with a centralized banking organization, performing in general the functions of a central bank, and to give the banking system as a whole a degree of unity and a coherence which it had previously lacked. As finally passed, the Federal Reserve Act represented a compromise between those who favored a single central bank and those who opposed any form of centralized banking. Because of the great size of the country, it was felt by some that regional banks would be better able to cope with the varying sectional problems than one large central bank, and so a system of twelve regional banks was adopted. The Federal Reserve did not supplant, but was superimposed upon, the existing banking structure.

At the head of the Federal Reserve System stands the Board of Governors of the Federal Reserve System, with headquarters in Washington. The board consists of seven members appointed by the President with the advice and consent of the Senate. They are

appointed for terms of fourteen years, with the terms so staggered that a vacancy occurs every two years.

In his selection of members of the Board, the President is required to give due regard to a fair representation of financial, agricultural, industrial, and commercial interests. To ensure representation on the Board for the various geographical districts of the country, the law provides that not more than one member may be appointed from any given federal reserve district. The President designates one of the members to serve as chairman of the Board of Governors for a term of four years.

Federal reserve districts and banks. The whole country is divided into twelve federal reserve districts, with a federal reserve bank located in a certain city in each dictrict. Each federal reserve bank is permitted to establish branches and agencies located in cities in its district and abroad. At present there are twenty-four branches and one agency. The accompanying map (Fig. 28 on page 320) shows the boundaries of the several districts and the location of the federal reserve banks and their branches.

A board of nine directors is entrusted with the management of each federal reserve bank. Six of the directors are chosen by the member banks, three to represent the interests of the member banks and three the interests of the business and industry of the district, while the remaining three directors are appointed by the Board of Governors. One of the latter is designated Federal Reserve Agent by the board and serves also as chairman of the board of directors of the bank.

Federal Advisory Council. The twelve federal reserve banks are given opportunity to participate formally in the making of the policies of the Federal Reserve System through the Federal Advisory Council. This council consists of twelve members, one selected annually by each federal reserve bank through its board of directors. It meets in Washington at least four times a year to confer with the Board of Governors on general business conditions and to make recommendations regarding the affairs of the Federal Reserve System. Its function is purely advisory.

Member banks. Each federal reserve bank is a corporation, and all its stock is owned by the individual banks within its district which have come into the system. These banks are known as "member banks." Each member bank's holding of federal reserve bank stock is equal to 3 per cent of its own capital and surplus. Every national

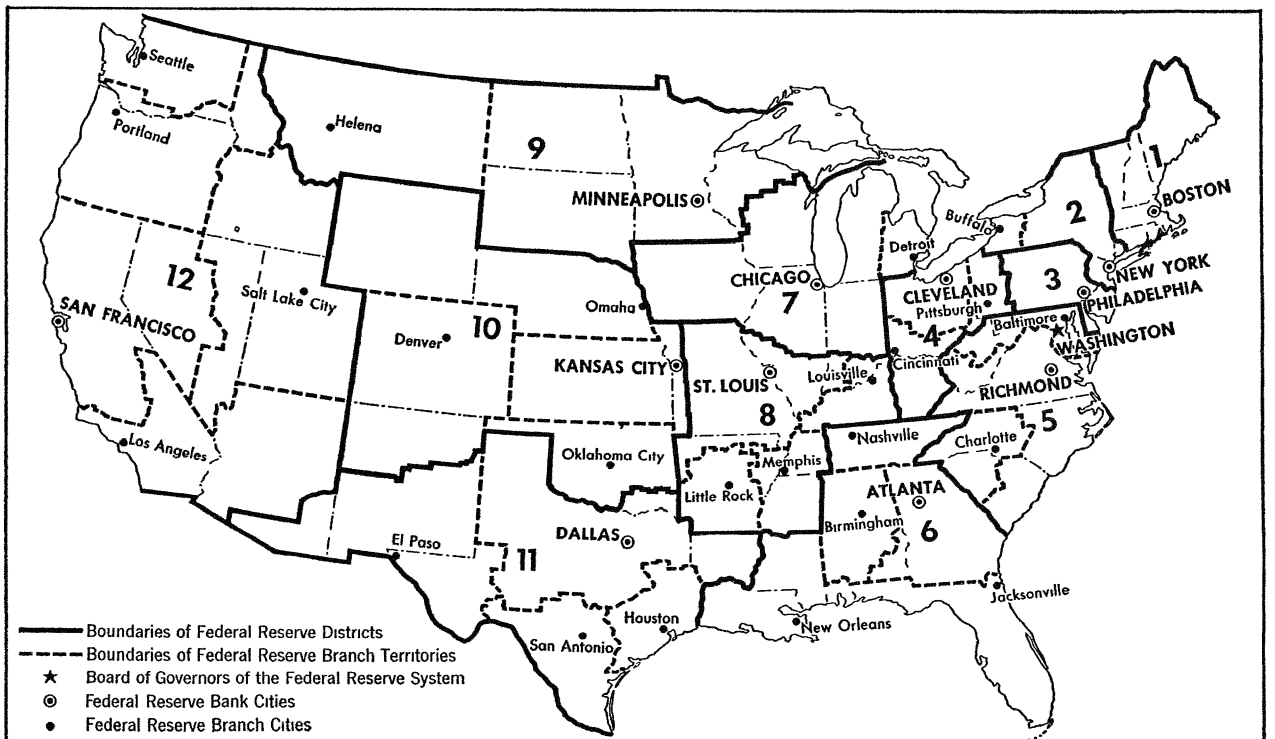


FIG. 28. FEDERAL RESERVE DISTRICTS AND BRANCH TERRITORIES

[1] Source: Board of Governors of the Federal Reserve System, *Federal Reserve Bulletin,* February, 1948. Washington: Government Printing Office, 1948.

bank is required by law to be a member, but membership is optional for the banks and trust companies incorporated under the laws of the states.

On December 31, 1947, there were 6,923 member banks, of which 5,005 were national banks and 1,918 state commercial banks and mutual savings banks. Only 20 per cent of the state banks were members, but these represented 47 per cent of the total loans and investments and 51 per cent of the total deposits of all state banks, showing that in general the largest and strongest institutions have come into the system. It is the small weak banks, most in need of the more stringent regulations of the Federal Reserve System, which still remain outside. In terms of resources, however, they constitute only a modest fraction of the nation's banking system.

Functions of the Federal Reserve: General. The banking business of the Federal Reserve System is conducted by the twelve federal reserve banks. They rediscount paper for their member banks and for each other; they hold deposits of the member banks; they issue notes; and they act to a certain extent as fiscal agents of the United States Government. They are permitted also certain dealings with the general public, but their principal business is with their respective member banks, with each other, and with the government.

Rediscounting and discounting. Rediscounting by the federal reserve banks is in harmony with the general principles with which the reader is familiar. The paper to be rediscounted by the member bank must be both eligible and acceptable to the federal reserve bank.

The basis for determining eligibility is laid down in the law, and in the regulations and rulings of the Board of Governors. Eligibility is based on the maturity and the purpose of the paper. Only short-term paper is eligible. When offered for rediscount, commercial paper may not be over three months from maturity, while for agricultural paper the maturity date must not be more than nine months distant. As to the purpose which gives rise to the paper, it is eligible if issued to finance the production, purchase, storage, or marketing of goods. Paper is ineligible for rediscount if the borrowers use their funds to speculate or trade in securities (other than those of the United States), or to speculate in commodities, or to purchase fixed assets.

Acceptability of paper is determined by (1) the soundness of the

paper itself, (2) the desirability of extending more credit to the particular bank, and (3) general business conditions and the probable effect of advances to the bank on general credit conditions.

A member bank may also procure federal reserve notes or other money from the federal reserve bank or increase its deposit account in the bank by getting so-called "advances"; that is, by discounting its own notes backed by government securities or eligible paper as collateral. Moreover provision has been made for member banks to borrow on noneligible paper. A reserve bank may make advances to member banks upon their demand or time notes having maturities of not over four months, secured to the satisfaction of the federal reserve banks. Such notes bear interest at not less than one-half of one per cent above the highest prevailing rediscount rate. Because of its greater cost, this type of advance is used only in case of necessity.

The discounting privilege has been further liberalized by a ruling of the Board of Governors in March, 1942, permitting reserve banks to make advances for periods not exceeding ninety days to both member and nonmember banks, on their promissory notes secured by direct obligations of the United States.

It was originally intended that dealings of the federal reserve banks with the general public be limited to open-market operations. Since 1932 amendments to the act permit federal reserve banks to do more extensive business with the public. They may now (1) make loans to individual firms, (2) discount for them obligations technically eligible for rediscount, and (3) make advances to such firms on the security of direct obligations of the United States Government. All this is to be done only when the individual firm is unable to receive adequate credit elsewhere.

Note issue: Federal reserve notes: General description. Two kinds of notes may be issued by the federal reserve banks: federal reserve notes and federal reserve bank notes.

The federal reserve notes, though nominally the obligation of the United States Government, are the direct obligation of the federal reserve bank of issue. They constitute, together with the federal reserve bank notes, a first lien against all the assets of the issuing bank. These notes are thus essentially bank notes of the normal type described in the preceding chapter, rather than government notes.

Security and reserve. Federal reserve notes are issued on the basis of collateral pledged as security with the Federal Reserve Agent. Initially each reserve bank was required to maintain a reserve in gold of not less than 40 per cent of its notes in actual circulation; in addition federal reserve notes could be issued only on the basis and to the amount of eligible paper pledged to secure them. Their notes were then backed by a 40 per cent gold reserve and in addition by a 100 per cent reserve in eligible commercial paper.

Eligible commercial paper was defined as short-term paper based upon commercial, industrial, or agricultural transactions. Thus the volume of federal reserve notes was to expand and contract with the needs of business. Federal reserve notes went into circulation when member banks rediscounted their customers' paper at the reserve bank, which in turn procured notes by pledging the discounted items with the Federal Reserve Agent.

Notes normally went out of circulation when customers repaid their indebtedness, for this enabled the member bank to liquidate its borrowings at the reserve bank and forced the reserve bank to retire an amount of notes equal to the repaid loans, unless indeed it had other paper to pledge with the reserve agent for that which had matured and been paid off.

Since the passage of the original act the definition of eligible collateral has been changed several times. At the present time a 25 per cent reserve consisting of gold certificates (gold having been withdrawn from circulation) must be held against outstanding federal reserve notes. Security to the amount of the remaining 75 per cent of the notes must be presented in the form of gold certificates or eligible collateral.

The paper collateral held as legal security against these notes may consist of the following: (1) notes, drafts, bills of exchange, and acceptances which have been discounted for individual firms by a federal reserve bank; (2) rediscounted eligible paper and collateral notes of member banks secured by eligible paper or government bonds; (3) bills of exchange; and (4) direct obligations of the United States.

On April 28, 1948, there were $24,473,210,000 of federal reserve notes outstanding, against which the federal reserve banks had pledged with the Federal Reserve Agents collateral security of

$13,159,000,000 in gold certificates, $85,094,000 in eligible paper, and United States Government securities amounting to $12,750,000,000. In other words, federal reserve notes are now secured almost entirely by United States Government securities and gold certificates, with the latter having no intrinsic value.

Loss of original concept of elasticity. As the Federal Reserve Act now stands, it constitutes a final and complete break with the intent of the original act, which was that federal reserve notes were to be issued to meet the needs of business as manifested by the discounting of short-term, self-liquidating commercial paper.

The founders of the Federal Reserve System contemplated that the federal reserve notes would ultimately constitute the largest part of the monetary circulation, and they believed that elasticity of note issue, with the volume of notes responsive alone to the seasonal, cyclical, and other changes in business needs, was assured by the previously described relation between the notes and eligible commercial paper. As matters have developed, while the federal reserve notes are as safe and sound as the credit of the government and the assets of the federal reserve banks can make them, the scheme of elasticity has been seriously upset by the substitution, in the main, of government bonds for the short-term, self-liquidating commercial paper as the basis for the issue of these notes.

But while the founders' plan for elasticity has thus gone askew, it must not be thought that the volume of federal reserve notes in circulation is invariable. The volume is affected by the need for hand-to-hand money, which varies with business activity, normally reaching a peak during the Christmas holidays. Any member bank can always meet an additional need for money by translating its deposits with the reserve bank into federal reserve notes. Moreover by the process of rediscount or borrowing on its collateral note, the member bank in sound condition can always increase its deposits with the reserve bank or its stock of federal reserve notes.

The local banks can meet the needs of their customers for currency in times of special business activity through granting deposits or giving federal reserve notes; when the special need is over, the currency contracts as loans are paid off, deposits decline, and the federal reserve notes are returned to the federal reserve banks in exchange for paper previously rediscounted and now paid off.

Federal reserve bank notes. End of the national bank notes. It was evidently the intent of the authors of the Federal Reserve Act

that the notes of the federal reserve banks should ultimately displace at least the greater part of the existing national bank notes. The law therefore repealed that part of the national banking act which required each national bank to invest a certain part of its capital in United States bonds. It further provided that, within certain limits, the national banks might sell to the federal reserve banks at par bonds which they had deposited to secure their notes and might retire the corresponding notes. The federal reserve banks, upon purchase of such bonds, were permitted to issue their own federal reserve bank notes secured by the bonds. As an alternative to issuing federal reserve bank notes, the federal reserve banks were permitted to exchange the 2 per cent bonds, thus bought from the banks, with the United States Treasury for new 3 per cent bonds or notes without the circulation privilege. As a matter of fact this program was never carried out; the interference of the First World War and the later reluctance of the national banks to surrender their circulation made it abortive.

The federal reserve bank notes have never become an important part of the monetary system. On June 12, 1945, Congress terminated the authority of federal reserve banks to issue such notes. There were $534,000,000 of federal reserve bank notes outstanding on June 30, 1945; by March 31, 1948, this amount had been reduced to $370,000,000.

In the year 1935, the Secretary of the Treasury called for redemption all outstanding bonds having the note-issue privilege. For this purpose the Treasury used the greater part of the gold "profit" from devaluing the gold dollar which was not devoted to the "stabilization fund." The entire issue of 2 per cent bonds, which for many years had been the only bonds to confer the circulation privilege, was thus wiped out by this action. The right to issue national bank notes against certain other United States obligations, which had been conferred by legislation of July 22, 1932, expired on July 22, 1935. In consequence, there existed no bonds eligible for securing national bank notes after August 1, 1935.

The Treasury has paid the national banks for their bonds and returned to them their redemption fund, and the banks have paid to the Treasury the sums necessary to redeem their notes. As fast as the national bank notes come into the Treasury, they are redeemed and canceled. This action results in some simplification of the currency system and reduces somewhat the advantage of the national

bank charter. On July 30, 1934, there were nearly a billion dollars in national bank notes outstanding. By March 31, 1948, the volume of national bank notes had been reduced to 102 millions, and doubtless much of this has been lost or has found its way into the hands of collectors.

Reserves against deposits: Legal requirements. The federal reserve banks receive ordinary deposits only from their member banks, foreign banks, and the United States Government and certain of its agencies; but deposits for the purpose of exchange and collection are received also from other federal reserve banks and from state banks and trust companies which are not members of the system. Against its deposits each federal reserve bank is required normally to keep a reserve of gold certificates (formerly gold) of at least 25 per cent.

Each member bank is required to keep with the reserve bank of its district a deposit not less than a certain percentage of its own deposits. The minimum *lawful reserves* are 3 per cent for time deposits; and, for demand deposits, 7 per cent for a country bank, 10 per cent for a reserve city bank, and 13 per cent for a central reserve city bank. For the purpose of this classification the main financial centers of New York and Chicago have been designated as central reserve cities. Certain other large cities have been designated as reserve cities, while banks in all the remaining cities and towns are designated country banks.

The Board of Governors has the authority to vary the reserve requirements provided it does not set them below the minimums just stated, or higher than twice those amounts. On July 1, 1948 the reserve requirements were 6 per cent for time deposits, and 14, 20, and 24 per cent respectively for demand deposits.

Centralized reserves. The Federal Reserve System is fairly well adapted to furnish a central reservoir and centralized control for the country's banking reserves. The only reserves against deposits now required by law of the member banks are their balances with the federal reserve banks.

The federal reserve banks have, on the other hand, maintained very large reserves, generally far in excess of the legal requirements. Their gold holdings were almost from the start extraordinarily swelled as a result of the suspension of gold payments throughout most of the world, the enormous demand for American exports, and other events of the First World War. In the wake of this war the

balance of exports in excess of imports remained large; moreover the political instability of Europe and other parts of the world inclined many people to try to preserve their fortunes by shifting the fluid part to America.

Mobile reserves. Not only do the federal reserve banks have a large reserve, in absolute amount as well as in relation to their liabilities and to the total monetary gold of the world, but this reserve, as well as the reserves of the member banks, is effectively mobilized. The banking reserves of the country are concentrated, available, and controlled in the public interest. Through dealings between the reserve banks themselves, directed and controlled by the Board of Governors, it is possible to direct the flow of reserves to the particular districts where for the time being the need is urgent.

Whereas in former days sound liquid banks were sometimes forced to fail because they could not obtain quickly the money due them from other banks, such an event is now virtually impossible. Any soundly managed member bank can convert resources into cash sufficient to meet any run, either by rediscounting with the federal reserve bank certain eligible paper or by borrowing on its notes with United States Government securities as collateral. On the other hand, a bank which has no eligible paper or government securities—a condition clearly indicating unsound management—has no means of rediscounting at the federal reserve bank though it may resort to the limit device of advances.[1]

Control of credit: Qualitative and quantitative control. The control of credit has both a qualitative and a quantitative aspect, and both have been provided for in the operations of the Federal Reserve System. The founders of the system placed chief emphasis on a qualitative control and sought by a strict and limited definition of eligible paper to confine the use of federal reserve credit to normal commercial, industrial, and agricultural short-term financing. For quantitative control reliance was placed on the regulation of the rediscount rate and on what are known as open-market operations.

In the development of the system, the forms of qualitative control have to some extent been elaborated, but through the liberalization of the definition of eligible paper, the effectiveness of this type of control has been considerably weakened. Quantitative con-

[1] See page 322.

trol moreover has been more strongly emphasized, and to the two original methods provided for such control has been added the power of the Federal Reserve to vary within limits the minimum reserve ratios of the member banks.

Control through rediscount rates. Just as the individual bank endeavors to control its reserve ratio by varying its rate of discount, so the power of the federal reserve banks to vary the rediscount rates charged member banks constitutes one of the available methods for controlling the volume of bank credit. By raising the rate the cost of borrowing at the federal reserve bank is increased, tending to dissuade member banks from making new loans. By thus slowing down the granting of new loans, while old loans are paid off at a predetermined rate, the volume of credit tends to be contracted. Conversely the lowering of the rediscount rate was supposed to expand bank credit by encouraging member banks to increase their own loans through increased borrowing at the federal reserve banks.

However the rediscount rate has not proved to be an effective instrument for quantitative credit control. The demand for short-term loans by business is influenced more by factors other than the interest rate and is therefore highly elastic. While normally changes in the rediscount rate have had their effect, as much for their indication of a change in central banking policy as for their effect upon the cost of loans, they were rendered virtually impotent during a period of heavy excess reserves, such as the late 1930's, and the consequent declining dependence of member banks for loans on the federal reserve banks.

The federal reserve banks have thus come to rely more heavily upon open-market operations, together with credit rationing and moral suasion, than on changes in the discount rate as methods of controlling credit. This stands in sharp contrast with the practices of the Bank of England, the Bank of France, and the other great central banks, whose influence over the national discount market is exercised largely through manipulation of their own discount rates.

Open-market operations. The dealings of federal reserve banks with the public, called *open-market operations,* consist of buying and selling in the public markets certain specified items—bills of exchange eligible for discount, bankers' acceptances, direct and guaranteed obligations of the United States and of its political sub-

divisions. Experience has taught that this is an effective instrumentality of control, permitting positive action by the authorities rather than passive waiting for requests from member banks for advances or rediscounts.

Purchases by federal reserve banks in the open market evidently tend to relieve tension in the money market, either by adding to the existing volume of money by payment for their purchases with federal reserve notes, or by increasing member banks' balances with the federal reserve banks, and thus augmenting the power of member banks to make loans and to expand their deposits. Similarly the sale of securities or of commercial paper on the open market will reduce either the amount of money in circulation or decrease member banks' balances with the reserve banks and so decrease the ability of member banks to expand their own deposit currency and perhaps even force them to contract it.

The increased importance attached to quantitative control of credit required that the open-market machinery be brought more fully and directly under the centralized authority of the Board of Governors. A Federal Open-Market Committee has been established consisting of the members of the Board of Governors of the Federal Reserve System and five representatives of the federal reserve banks. The committee adopts and transmits to the federal reserve banks regulations relating to the open-market transactions of such banks. The open-market operations of every federal reserve bank must be in accordance with the direction and regulations of this committee. The time, character, and volume of all transactions in paper eligible for open-market operations must be governed with a view to accommodating commerce and business, and with regard to the bearing of such transactions upon the general credit situation of the country.

Open-market operations can be successful as a means of expanding member bank reserves because of the tremendous quantity of eligible paper available for purchase by the federal reserve banks. However, mere expansion of reserves is no guarantee that the member banks will expand their loans. For in the last analysis, the quantity of loans depends upon the demand for loans, and the Federal Reserve System has no direct means of increasing this demand.

On the other hand, open-market operations are not always successful as a means of contracting credit. Federal reserve banks may

sell securities on the market in the attempt to reduce member banks' reserves and thus force them to restrict their loans, but such a program is successful only when member banks have small reserves. Because of heavy gold imports and other factors, member banks had large excess reserves all during the 1930's. Even had the reserve banks sold all of their securities, member banks' reserves would still have been substantially in excess of requirements.

During the Second World War the situation was radically changed; member banks had relatively small amounts of excess reserves, and the Federal Open-Market Committee pursued a policy of easy money by standing ready to purchase all eligible paper at any time.

Regulation of reserve ratios of member banks. The most recently developed instrument of quantitative control of credit is the regulation of the member banks' reserve ratios, *i.e.,* the percentages of deposits which they must hold as reserve with the federal reserve banks. Under the banking act of 1935 the Board of Governors of the Federal Reserve System may, "in order to prevent injurious credit expansion or contraction," change by regulation the percentage requirements. If a bank required to carry a 10 per cent reserve were directed to jump its reserve to 20 per cent, its loaning capacity would be reduced, and a lowering of the reserve percentage later would enhance this capacity. Of course, instead of curtailing loans when the reserve requirement is increased, the bank may sell securities from its portfolio and transfer the proceeds to its reserve account.

Nevertheless this method of control is even more direct and powerful than open-market operations in creating or absorbing excess reserves. Moreover by reducing reserves and forcing members to become indebted to the federal reserve banks, the board can make changes in the rediscount rate more effective in credit regulation. Such control is a new device in central bank technique; how useful it may prove to be is problematical.

A national system of clearing. The Federal Reserve System has also become a national clearing house for settling the claims of the reserve banks against each other and for collection of checks, drafts, and other claims of individual banks against out-of-town banks, both in its own and in the other federal reserve districts. While the system is far from universal, on December 31, 1947, there were

12,071 banks,[1] 86 per cent of the total, which had agreed to remit at par for checks sent by the federal reserve banks for payment.

Government fiscal agency. The reader will recall that, following the banking collapse in 1837, Congress established the Independent Treasury system, and that later a variable number of national banks were used as depositories of public funds. The Federal Reserve Act partly removed the distinction in favor of national banks by extending the privilege of holding government deposits to state banks that were members of the Federal Reserve. Of greater significance was the provision that the funds of the United States might be deposited with federal reserve banks and that these banks might be required to act as fiscal agents of the Treasury. In 1921 the subtreasuries were abolished and the Treasury has made progress towards curtailing its reliance upon the local banks and concentrating its fiscal business in the hands of the federal reserve banks.

Today the twelve federal reserve banks carry the principal checking accounts of the United States Treasury, receiving payments from taxpayers and purchasers of government securities, which are credited to the Treasury's account, and paying the government's checks, which are charged to the Treasury's account. The reserve banks handle most of the work of issuing and redeeming government obligations, and act as custodians of collateral and securities in connection with the lending and other financial activities of many government agencies, such as the Reconstruction Finance Corporation. Finally the Federal Reserve Bank of New York acts as the agent of the United States Treasury in the foreign exchange operations of the Treasury's Stabilization Fund.

[1] Represents banks on which checks are drawn, except that it excludes both member and nonmember mutual savings banks, on a few of which some checks are drawn.

CHAPTER XVII

Recent Developments in American Banking

A decentralized banking system: The local banks in 1929. Although the Federal Reserve brought to American banking more semblance of order and centralized control than it had had before, the local banking system still remained what it always had been, a multitude of small independent local banks. This condition has persisted down to the present time.

For its ordinary banking needs the public is served by (1) the national banks, under federal charter; (2) state banks, trust companies, and savings banks, chartered by the several states; and (3) unincorported private banks. It will be recalled that all national banks must be members of the Federal Reserve System, while for the state institutions membership is optional. The accompanying

BANKING SYSTEM OF THE UNITED STATES, JUNE 30, 1921 AND 1929

I. All Banks

	NUMBER		RESOURCES (IN BILLIONS)	
	1921	1929	1921	1929
National banks	8,154	7,536	$20.5	$27.4
State banks	18,875	14,437	14.2	16.8
Loan and trust companies	1,474	1,608	8.2	16.1
Stock savings banks	978	747	0.5	1.5
Mutual savings banks	623	611	6.0	10.0
Private banks	708	391	0.2	0.2
Total	30,812	25,330	$49.6	$72.0

II. Federal Reserve Membership

	NUMBER		RESOURCES (IN BILLIONS)	
	1921	1929	1921	1929
National banks	8,154	7,536	$20.5	$27.4
State member banks	1,595	1,171	10.0	18.5
Total member banks	9,749	8,707	$30.5	$45.9
Nonmember banks	21,063	16,623	$19.1	$26.1

table will serve to give a broad statistical picture of this system as it was in 1921 and in 1929.

Omitting private banks and savings banks, the country's banking business was handled in 1929 by something over 23,500 banks. About a third were national banks, having average resources of about $3,500,000 and owning slightly over 40 per cent of total bank resources, while the remainder were state banks and trust companies with average resources of around $2,000,000. This is in marked contrast with the situation in most other countries; in Canada, for example, there were in 1928 only 10 large central banks, with average resources of nearly $350,000,000.

Even these figures do not show as convincingly as do the figures on bank capitalization the extent to which the small bank prevailed. In 1929 only 28 per cent of the banks of the country had capital of $100,000 or over; of the remaining 72 per cent, 21.9 per cent had less than $25,000, 21.5 per cent had exactly $25,000, and the rest between $25,000 and $100,000. Over 80 per cent of the banks were located in towns of 10,000 or less population, but these banks accounted for only 23.4 per cent of the total capitalization.

Branch banking limited. Without doubt the characteristic decentralization of banking in the United States would not have persisted as it has, had branch banking been generally permitted. Only in recent years, however, have there been any important steps away from the prevailing condition of local decentralization. In California and certain other states, the state law has for some time permitted branch banking, and great systems of branches have been developed, state-wide, county-wide, or city-wide.

The competition of these state banking institutions finally forced Congress, after a long and bitter legislative struggle, to grant to the national banks certain branch banking powers. The McFadden act of 1927 provided that in any state which permitted branch banking both national and state members of the Federal Reserve System might, with certain limitations as to national banks, establish and maintain branches within the limits of the city where the parent bank was located. By 1929 there were 818 banks operating 3,440 branches in the United States, with the great majority of the branch offices situated in the local communities of the respective parent banks. While these banks with branches formed an important part of the banking system, owning about 40 per cent of the loans and investments of all banks, branch banking had not proceeded ex-

tensively enough to alter the essential character of the American local banking system.

Another development of considerable importance has been the establishment of elaborate "chains" or "groups" of banks, held together by various devices, the most common being the holding company. This was a device which in fact achieved branch banking, but without the sanction of state or federal law. These bank-holding companies were not subject to federal control and supervision until made so by the banking act of 1933.

In June, 1929, there were 275 such chains or groups of banks in the United States, controlling 1,802 unit banks, which in turn owned 40 per cent of all the branches in the country. The loans and investments of all banks controlled by chains or groups amounted to about 15 per cent of the loans and investments of all banks. Nevertheless important as they were, the chains did not permeate the banking structure sufficiently to be a major factor in the situation.

Weaknesses of decentralized banking. There are serious inherent weaknesses in a system of small independent local banks. By the multiplication of banks, the country's banking capital and deposits are so minutely divided that many local banks are too small to provide effective management or service.

Banking had in 1929 become one of the most overexpanded industries. The small capital required under state and federal laws made it easy to organize a bank. Towns were jealous of one another, and so each had to have its bank. Dissatisfied customers of one bank or ambitious local capitalists or adventurers would often start a second or a third bank in a town which could not decently support one.

Competition among these banks, many in a failing condition, was severe, ruthless, and unintelligent. High interest rates were paid on deposits, and many services were performed without charge and at a loss. The banks competed in "service to the community," which largely meant buying local securities and making loans on hope and uninformed faith rather than upon an intelligent analysis of the borrowers' ability to repay. Thus each bank's portfolio of loans and investments contained too many local items, and the life of the bank became dependent upon the stability, permanence, and success of the small locality.

This competition had effects on banking practices. By the sheer

necessity of acquiring sufficient earnings on a relatively small volume of loans to support their overhead costs, the banks were frequently forced to charge exorbitant interest rates for loans, and at such rates loans could be made only to necessitous local borrowers. Such investments as could be made in securities traded on the open market had for the same reason to be made in nominally high-yield bonds and mortgages regardless of safety or liquidity. Bank buildings were frequently excessively elaborate in architecture and equipment, and the size of the staff was excessive. Salaries, however, were often too small to be attractive to men of adequate training and capacity.

The small-town state banks were on the whole indifferent, if not hostile, to the Federal Reserve System. In 1929 only 7 per cent of all the state banks belonged to the system, and these were the larger banks. The national banks in small towns or cities had a difficult time competing with the state banks, for in many states the banking laws were more liberal and the supervision more lax. This made the way of the national bank supervisors difficult, since severity was frequently answered by threats to denationalize and take out a state charter. The difficulty was partly adjusted by lowering the national bank standards (by the act of 1927 and other amendments) as well as by less punctilious examinations.

Situation in the decade prior to 1933: Expansion of credit. Among the financial features of this period which had a bearing upon the banking situation in 1929, we note first a gigantic expansion of credit in the United States, marked by an increase of about fifteen billion dollars in loans and investments of commercial banks, with the total for all member banks of the Federal Reserve System rising to thirty-six billion dollars. Such an expansion was made possible by the inflow of gold from countries off the gold standard or on a gold-exchange standard and by the "easy money" policies of the federal reserve banks. This federal reserve policy was dictated partly by political and industrial conditions in the United States and partly by a program of coöperation between the Federal Reserve Bank of New York and the Bank of England. Most of this increase in credit was neither needed nor used for normal short-term commercial operations but went into speculation in real estate, speculative construction, and the taking up of a tremendous issue of securities, both domestic and foreign.

Lending and investing abroad was a rather novel practice for

Americans, but at the end of 1932, after considerable reductions from both default and resale, the private foreign investments of the United States amounted, conservatively estimated, to fifteen billion dollars, of which seven billions were represented by security investments in the ordinary sense and eight billions by direct ownership of capital assets.

The annual lending of immense sums abroad had permitted the merchandise export balance of the United States to expand, in spite of rising trade barriers and against the influence of net payments due the United States because of its creditor position. This made possible large exports of heavy mechanical equipment and of agricultural products, which served to maintain these types of enterprise in precarious balance, despite the overexpanded condition of their productive capacities in relation to the domestic market and in relation to the reviving and expanding productive capacities of other parts of the world.

The new annual loans also served to conceal the inherent difficulties of payments on the intergovernmental debts and reparations in a world of rising trade barriers and badly distributed monetary stocks of gold.

Financial crises in 1929 and 1931. Despite persistent depression in agriculture, industry in the United States had with minor exceptions been fairly prosperous after 1920, but in 1929 the tide turned. Funds in the United States were increasingly in demand for speculation in the domestic security markets. It was becoming evident that the structure of international finance was unsound. With foreign lending reduced, the export industries of the United States were either losing their foreign markets or saving some of them by taking new credit risks and selling on open account, and the unnatural export balance of the United States was declining.

By the summer of 1929 most lines of business activity in the United States were showing some recession, and this was fatal to a domestic structure of speculative credit which was based on expectation of everlasting expansion, of accelerated profit taking, and of security values which could move only up. Liquidation made a sudden start on its predestined long journey, foreign lending came to a standstill, and the postponed and now intensified problems of postwar economic readjustments presented themselves for fresh consideration.

In the summer of 1931 a crisis was reached in international

financial relations, with successive breakdowns in Austria, Germany, and Great Britain; this was despite considerable international coöperation, in which the United States participated by moratoria on the allied war debts and on commercial debts, as well as by special advances to foreign central banks. Great Britain left the gold standard in September, 1931, and was followed by upwards of thirty countries, leaving a slender "gold bloc" consisting of France, Holland, Belgium, Switzerland, and Poland. There was universal withdrawal of balances of short-term capital from foreign banks, especially from the United States. Among the consequences of this action were renewed credit liquidation, heavy cash withdrawals from the banks, and increased failures in the United States.

A decade of bank failures. From 1921 through 1932, 10,816 banks suspended; 5,102 of these failures occurred within the three years, 1930, 1931, and 1932, following the crash of 1929. This is the worst record of bank failures in the world's history, despite the fact that the United States was the wealthiest nation in the world and had gained most and suffered least from the First World War. With a sound banking system this crash would have been unnecessary. Neither England nor Canada was completely insulated from the influence of the forces at work in the world at that time, yet not a single bank in England or in Canada failed because of the depression.

Analysis of these bank suspensions showed that the rate of failure was higher among nonmember than member banks, among banks in the smaller ranges of capitalization, among banks in the towns and cities of small population, among banks in agricultural areas as compared with industrial territory, and among state banks as compared with national banks because of their smaller size and looser supervision.

In the early months of 1933 the American banking system virtually broke down. Among the causal events leading up to this crisis, the following are especially deserving of mention: accumulation of huge deficits in government financing, a substantial decline in commodity prices, bold discussion in Congress of various fiat (irredeemable) money proposals, the suspension between 1929 and July, 1932, of approximately a fifth of all the banks operating in the earlier year, further increase of bank failures after November, 1932, loss of gold by export, and an unprecedented popular movement toward the hoarding of gold.

In January and February, 1933, there were mass closings of banks by cities and states. Finally on March 5 came the proclamation of President Roosevelt closing all the banks in the country for a period of four days beginning on March 6 and placing an embargo on gold withdrawals and exports. Congress was called in special session on March 9 and on that day passed the emergency bank act, giving the President emergency powers over the banking system.

Structural changes since 1929: Introduction. Congress and the state legislatures have since 1929 enacted many important laws to mend the weaknesses of the banking system which were demonstrated by the stock market boom of 1926–29, the panics of 1929, 1931, and 1933, and the depression of 1929–36. This legislation, especially that of the federal government, together with the effects of the adjustments made to these successive crises, has changed considerably the structure of the banking system and has put into effect new theoretical conceptions of the nature and functions of central banks.

Branch banking. Branch banking received modest encouragement in the banking act of 1933 by the grant to national banks of authority to establish and operate branches, either in the city or throughout the state in which the bank is located, provided that similar authority to maintain branches is specifically granted to state banks by the laws of the state in question. This permission is, however, subject to severe restrictions as to the population of the place and the size of the bank, which are intended to limit the establishment of branches in small places and by small banks. A number of states have liberalized their banking laws in consequence of this action, but even today branch banking is still operating on a comparatively small scale.

Branch banking offers one important means of correcting the faults of the decentralized local banking system. It affords greater safety by permitting a wider diversification of loans and investments. In addition, because of the less intimate nature of the branch bank, the granting of loans can be made on the basis of a more scientific credit appraisal. Deposits are also safer because they are spread all over the nation, though still backed by the aggregate resources of the parent bank. Capital is more mobile, and interest rates tend to be more uniform. The greater diversification, as well as the ease in shifting funds to troubled areas, makes a branch banking system more capable of resisting the shock of depressions.

The general level of service is higher, since each branch enjoys the use of all the specialized services offered by the parent bank and can command better managerial skill. Finally banking facilities can be more numerous without suffering from the danger of undercapitalization that faces small unit banks.

Holding company affiliates. The banking act of 1933 gave some attention to the related problem of chain banking—the problem arising from the control and direction of a group of banks by a holding company. The act requires such a holding company, called a holding company affiliate, to secure a permit from the Board of Governors before it may vote the stock which it owns in a member bank. To obtain the voting permit the holding company must submit to examination like any national bank, must build up assets other than bank stock, and must dissolve any connection with a security selling organization. Member banks are restricted in the making of loans to a holding company affiliate, in the purchase of its obligations, and in the acceptance of such obligations as collateral for loans to others. Of any group of member banks affiliated with a holding company only one may participate in the nomination and election of directors of a federal reserve bank. Some measure of control was thus extended over a form of banking organization which had been open to considerable abuse.

In recent years, the place of bank chains or groups in the American banking scene has been steadily declining in significance. Thus as of December 31, 1943, there were in existence only 131 groups and chains, controlling 837 unit banks, which were operating but 22 per cent of all the branch banks in the United States. Furthermore, the combined deposits of these chains were only $15,266,000,000, being 14 per cent of the total deposits of all commercial banks. This gradual decline of chain banking may be attributed in part to the greater efficiency and desirability of branch banking as a means of accomplishing the same purpose.

Capital requirements. As a step toward the elimination of difficulties arising out of the keen competition in a banking system composed of very many small and frequently undercapitalized independent banks, recent legislation has altered capital requirements.

Legislation in 1931 and 1935 increased the minimum capital for national banks to $50,000 for a bank in a place having a population of less than 6,000, with correspondingly larger requirements in

the larger localities. State banks are excluded from membership in the Federal Reserve unless the national bank capital requirements are met.

The banking act of 1935 further requires that newly organized national banks must have surpluses equal to 20 per cent of their capital stock, and that every national bank must build up its surplus out of earnings to a level of 100 per cent of capital stock instead of the 20 per cent level formerly required. In one respect this accumulation of larger surplus is offset by the loss of the security afforded to depositors by the double liability of shareholders, which has now been abandoned.

Divorce of commercial and investment banking. The reader has been reminded that the keenly competitive search for banking profits during the 1920's developed unwise policies in the granting of credit by commercial and investment banks. Commercial banks devoted more and more of their resources to strictly speculative loans, to real estate loans, to investment holdings, and even to dealings in securities. The shift of funds away from strictly commercial banking functions was intensified under the conditions of business depression.

In fact there were instances of close affiliation of commercial banking with other forms of enterprise. Not infrequently close relationships between commercial banks and companies dealing in securities or in real estate were created by the holding company device; in other cases the bank itself owned the controlling stock in an affiliated company.

It is the general consensus of opinion that a commercial bank should be divorced from the investment business, real estate promotion, and other such interests, especially if it has a pecuniary interest in the profits of the other enterprises. Indeed it is difficult to see how an officer of a bank can bring an unbiased judgment to bear on the wisdom of making a loan to a customer which will enable the latter to purchase an issue of stock or a piece of real estate offered for sale by a company affiliated with the bank.

The banking act of 1933 sought to correct these defects. Limits were placed on loans by any member bank to its own officers, loans to security affiliates were forbidden, and the amount of the obligations of such affiliates which might be accepted as collateral on loans was restricted.

Another group of provisions was intended to effect a divorce be-

tween commercial and investment banking. Member banks were forbidden to engage in the investment banking business; a limit was placed on the amount of any one issue of securities that a member bank might hold on its own investment account; and it was stipulated that no member bank might be affiliated in any way with an organization engaged principally in the issue or distribution of securities, and that no officer or director of a member bank might be an officer, director, or manager of any organization engaged primarily in the business of buying, selling, or negotiating securities, unless with permission of the Board of Governors. On the other side, no enterprise engaged in investment banking was permitted to engage in the business of receiving deposits. The banking act of 1935 slightly relaxed some of these prohibitions and gave the banks a limited privilege of underwriting and dealing in securities.

Centralized control of the banking system. The Board of Governors had not, up to 1933, attained to that dominance in the banking system which was regarded by many as desirable. A number of changes were accordingly made, designed to bring about greater integration within the Federal Reserve System itself through a strengthening of the control of the Board of Governors. This control was fortified by the power granted the board to alter reserve requirements, to regulate the payment of interest on time and savings deposits, to exercise supervision over credit for speculative purposes, to establish regulations governing the standards for rediscounts and loans by federal reserve banks to members, to set the rediscount rates of the banks, and to control through the Open-Market Committee all open-market operations and impose its decisions upon the federal reserve banks. The control was further strengthened by the power granted in 1933 to order the removal of any director or officer of a member bank if, after warning, he continued to violate any banking law or continued "unsafe or unsound practices."

Government banking: Reconstruction Finance Corporation. The increasingly critical condition of American banks, evidenced by the failure of 1,072 banks in the five months beginning with September, 1931, and the intensification of hoarding, led the United States Government to take various measures to support the credit situation. The first of these was the organization in October, 1931, of the National Credit Corporation for self-help among the banks. This involved the pooling of resources so that loans might be made on

call to distressed members. The National Credit Corporation did not, however, prove to be of much service, and in February, 1932, it was replaced by the newly created Reconstruction Finance Corporation.

The Reconstruction Finance Corporation was originally established by act of Congress as an emergency agency to lend funds to distressed financial institutions, railroads, and farmers. Its life has been extended several times, and today it is a gigantic corporation, in reality a huge government bank, with thirty-one branches throughout the country, playing a predominant role in both long-term and short-term finance. The corporation has become intimately articulated with the Treasury and the financial system of the country.

The RFC started with capital stock—at present, 375 million dollars—subscribed by the government, and was authorized to borrow on its own obligations up to three times the amount of its capital. It was empowered to make loans to banks, building and loan associations, insurance companies, mortgage companies, federal agricultural banks, agricultural and livestock credit corporations, railroads, and the farmers. Later the lending power of the corporation was broadened to permit loans for self-liquidating public works undertaken by states, municipalities, and other public or semi-public agencies. The central purpose of the corporation was to use government credit for thawing frozen private credit.

The RFC has participated in many large and important credit operations of the government since its formation. It provided the funds to reopen the banks in 1933 and to prepare them for deposit insurance, as well as to finance the liquidation of failed banks and the establishment of many of the emergency governmental agencies, such as the Home Owners Loan Corporation, the Export-Import Bank, and the Rural Electrification Administration.

In 1933 the corporation initiated an emergency measure which later became a fixed policy for handling emergencies; *i.e.*, the practice of making loans at first indirectly through the banks and later directly to business firms themselves to provide them with working capital. These latter loans authorized by law in June, 1934, were permissible only in the event that industrial firms were "unable to obtain requisite financial assistance on a reasonable basis from the usual sources."

From the beginning of the defense program in 1940 through to

the postwar reconversion period, the RFC has discharged such essential banking functions as financing activities essential to the national defense; industrial war plant conversion and construction, "stockpiling" strategic war materials, insuring property against losses from enemy attack, disposition of capital and producers' surplus war assets, and the making of loans to small business concerns.

These extensive credit functions put the RFC in direct competition with the federal reserve banks in rediscounting the industrial paper of individual banks, and with the banks themselves in making direct loans to business concerns. That this competition was not inconsequential is shown by the fact that up to March 31, 1944, the RFC had disbursed 25½ billion dollars, of which 22 billions were loans. However, the corporation was established to meet the unprecedented conditions of the business emergency when private and normal banking credit was largely frozen. Its powers to extend credit were purposely made liberal so that it could reinforce with government credit the shrinking credit of private sources and thereby stimulate business recovery.

Agricultural credit agencies. In the special banking field of agricultural credit, the federal government has assumed new or amplified banking functions. The various agricultural agencies were consolidated into a new Farm Credit Administration under a single Governor, for the purpose of improving the efficiency of the credit service. The administration, in addition to the Governor, consists of the Land Bank Commissioner and the Intermediate Credit Commissioner, each of whom heads a centralized system of credit institutions. The institutions thus consolidated consist of (1) the twelve federal land banks—long-term mortgage banks, owned by some 3,000 coöperative national farm loan associations scattered over the country; (2) the joint stock land banks—long-term mortgage banks, similar to the federal land banks, but privately owned and operated and to be liquidated, no new loans being permitted; (3) the twelve federal intermediate credit banks—operating in the same districts, housed in the same buildings, and governed by the same directorate as the federal land banks, but lending for periods from one to three years on agricultural products and livestock; (4) twelve production credit corporations, supplied with government funds and authorized to invest in the stock of, make loans to, and in large measure control the activities of, local production credit associations, which lend to farmers for general agricultural

purposes; and (5) the Central Bank for Coöperatives and twelve regional banks for coöperatives, with funds largely supplied by the government, authorized to make loans to coöperative associations which promote the orderly marketing of farm crops by holding, storing, processing, and marketing farm products, and by making loans to their members on the security of farm products deposited with them for eventual marketing.

The federal government is thus prepared to perform functions which in times past were considered to be exclusively within the province of the banks, but which were not fully performed by them during the long period of bank closings and liquidations in agricultural areas. The federal farm agencies offer serious competition to commercial banks. They have entered into all the lending activities of agricultural areas, and furthermore they charge a lower rate of interest.

Congress went still further and created the farm mortgage corporations, to issue and use two billions of dollars of refinancing bonds, guaranteed by the government as to both principal and interest. The funds of the corporations are used (1) to refund bonds which had already been issued by the federal land banks for the purpose of refinancing farm mortgages, thus exchanging new bonds fully guaranteed for the former bonds, which were guaranteed only as to interest, and (2) to refinance mortgages of additional mortgagors who might be induced to exchange mortgages for the newer type of bonds.

Housing credit agencies. The home loan mortgage business had never been adequately provided with funds or effective machinery for granting, servicing, or rediscounting loans. One significant effect of the failures, in 1929 and thereafter, of the state and national commercial banks, as well as of the mortgage bond and mortgage guarantee companies, following as they did the excessive real estate developments between 1922 and 1929, was the complete prostration of the construction industry and the collapse of real estate values. It was felt by some that recovery from the depression would be impossible unless this situation were relieved and the construction industry were revived. Into this situation stepped the government with a series of credit institutions.

These agencies act as a directing and supervising organization and are themselves organized, in addition to the Office of the Administrator, into three constituent units: (1) the Federal Home

Loan Bank Administration, (2) the Federal Housing Administration, and (3) the Federal Public Housing Authority.

The Federal Home Loan Bank Administration supervises a group of agencies which (1) extend to mortgage companies and the federal savings and loan associations help in financing their lending operations, especially in emergencies, (2) do business similar to that of the building and loan associations, (3) insure the shares in the federal savings and loan associations and other members of the federal home loan banks, and (4) from 1933 to 1936 purchased with government funds mortgages on homes and made supplemental cash loans when these were needed for such items as accrued taxes and repairs.

The function of the Federal Housing Administration is to insure private lending institutions against loss of loans secured by mortgages on one- to four-family dwellings or large-scale rental projects, and on loans for property repair or improvement.

The Federal Public Housing Authority provided housing for persons engaged in war activities in areas where acute shortages existed, and it now assists local public housing agencies in low-rent housing and slum clearance projects.

The National Housing Act also provided for the establishment of national mortgage associations, under federal charter, with power to buy and sell first mortgages and to borrow money for such purposes by the issuance of securities. One such institution, the Federal National Mortgage Association, of Washington, D. C., was created by the RFC in 1938, but to date no other national mortgage association has been established.

Criticism of government banking. To a certain extent specialized banking activities of the government may be regarded as necessary salvage operations in an emergency characterized by general timidity and frequently paralyzing losses among private credit agencies. A great deal of hardship for the families of farmers and of urban home owners was averted or ameliorated. At the same time it must be recognized that putting government capital into competition with private capital on an extended scale does not tend to restore the active confidence of the latter in the prospects for profitable operations. The history of the assumption of functions by government suggests the possibility that much of the operation of specialized banking services by the government of the

United States may be destined to continue indefinitely, a prospect which is scarcely calculated to inspire confidence on the part of private lending agencies.

Functional changes since 1929: Control of interest on deposits. Of the changes instituted in the functions of banks by recent legislation, considerable importance attaches to the institution of central bank regulation of interest rates on deposits. The acts of 1933 and 1935 prohibit member banks from paying interest on demand deposits, and direct the Board of Governors of the Federal Reserve System to limit by regulation the rate of interest paid by member banks on time and savings deposits. In addition the Federal Deposit Insurance Corporation must by regulation prohibit the payment of interest on demand deposits in insured nonmember banks and similarly must limit from time to time the rates of interest paid on time and savings deposits. The corporation and the Board of Governors have coöperated in fixing the same rates in any territory, and they have also secured the coöperation of the state banking departments in setting rates for noninsured banks.

This change in the operation of the banking business was directed at one of the difficulties of decentralized banking. The interest allowed on deposits was frequently a chief competitive weapon among banks in their attempts to secure an adequate volume of business to provide a fair return on capital. There is no basic need for interest on the demand deposits of a commercial bank, and such payments merely encourage the keeping of unnecessarily large balances in a form which is most dangerous for the bank.

The prohibition of interest payments on demand deposits was more or less welcomed by the banks at this time; first, because it offset in part the cost of premiums for federal deposit insurance; second, because the banks had found it very difficult to earn their costs and dividends during the depression while safe loan opportunities had been so few and yields on investments so low; and third, because they had been surcharged with excess funds held idle for the sake of liquidity.

Deposit insurance: Description of the federal system. With the purpose of allaying the public distrust of banks which the banking crisis of 1932-33 may have generated, and serving as a preventative of similar panics in the future, a nationwide system of deposit insurance was established. The system does not provide government

guaranty but rather self-insurance of their deposits by the banks themselves. The Federal Deposit Insurance Corporation administers the program.

Participation in the plan is required of member banks of the Federal Reserve System and is optional with nonmember banks. Through a compromise between the advocates of government insurance and the proponents of decentralized banking, the law originally required the largest of the state banks—with average annual deposits of over a million dollars—to become members of the Federal Reserve System and thus automatically enter the deposit insurance plan on or before 1942. But in 1939 the law was amended to permit all state banks to remain nonmember banks indefinitely and still at their option participate in the deposit insurance.

The FDIC may, after notice, suspend an insured bank from the insurance system and publicize its action if the bank continues unsound or unsafe practices or knowingly or negligently permits violation of the law or regulations. Any insured nonmember bank, upon due notice, may voluntarily withdraw from the insurance plan.

The insurance covers a maximum amount of $5,000 for any depositor in one bank; it is financed by an assessment on all participating commercial banks amounting to one-twelfth of one per cent per annum of average net deposits.

The FDIC was authorized at its discretion to operate a separate fund for the benefit of the mutual savings banks and their depositors. Conforming to the authority, such a fund is now operated in a wholly separate manner for the mutuals alone, and the capital assets of the corporation, as well as all its expenses, are allocated between the two funds on an equitable basis. Such separate treatment of mutual savings banks was warranted by the fact that their deposits are all time deposits and that the mutuals operate, as experience shows, with a very low loss ratio. The scheme has not, however, been popular with the mutuals. The corporation in 1947 was insuring the deposits of only 194 mutual savings banks out of a total of 533.

The extent to which deposits are now insured is impressive. Of the 161,850 million dollars deposited in the 14,714 banks in the United States on December 31, 1947, only 7,792 million dollars were in noninsured banks. Although well over 95 per cent therefore of all deposits were in insured banks, less than half of these

deposits were covered by insurance because of the $5,000 limit. Of the 19,046 banks, branches, and additional banking offices operating at the end of 1947, 17,810 were insured.

Appraisal of deposit insurance. The experience in the United States with various schemes for the guaranty of bank deposits would seem to warrant misgivings as to the right of a deposit insurance plan to a permanent place in the banking structure, however essential it may have been considered in 1933 as a means of overcoming the quite natural timidity of depositors and in counteracting the trend towards hoarding. A number of states in the United States have experimented with guaranty of bank deposits, but invariably the standards in banking practice were lowered. In every case the plan quickly broke down and was abandoned. The continued operation of the schemes would have put all solvent banks in bankruptcy.

A national scheme of deposit insurance undoubtedly stands a better chance of success than the state schemes. For one thing, there is much greater diversification of risk than in a state scheme, where all the banks are frequently dependent on a few types of loan and in danger of all getting into difficulty together if any economic adversity strikes the leading industry of the region. Again it may be expected that the supervision of the banks under the Federal Deposit Insurance Corporation will be stricter than that which existed in most of the states. The corporation has, in fact, already set higher standards for admission to the system. In times of severe depression, the federal government is better able to come to the assistance of the insurance fund if it runs short of reserves.

It cannot be denied that the federal deposit insurance scheme strikes at the effect rather than the cause. It perpetuates the present banking structure, with all its weaknesses. Undoubtedly it would be desirable to correct these faults rather than to cover them up by an insurance system. But until these faults have been eradicated—a condition which may be many years in coming—it may be better to install deposit insurance in the hope of avoiding the dire effects that frequently accompany bank failures.

The jurisdiction of the Federal Deposit Insurance Corporation offers of course a new way to exercise a centralized influence over most of the 7,791 state banks which are not members of the Federal Reserve System. So long as they feel the need of deposit insurance, these banks, even though they fail to join the Federal Reserve, are

actually subject to many of the regulations of that system. The possible centralizing influence is further strengthened by the opening of membership in the Federal Reserve System to Morris Plan banks and others of the same type, and to mutual savings banks that qualify.

Loans on real estate. Loss of interest in qualitative control of credit is further indicated by a provision of the banking act of 1935 which greatly extends the authority of national banks to loan on real estate security. This represents a departure from the notion of keeping these banks strictly commercial.

A national bank is now authorized to make loans secured by first liens upon improved real estate in amounts not exceeding 50 per cent of the value of the real estate; the maximum period of a loan is five years, except for amortized loans, which may be made in amounts not exceeding 60 per cent of the appraised value and may run for ten years if the installments are such as will amortize at least 40 per cent of the principal of the loan within the ten-year period. These restrictions do not apply to insured real estate loans made under the provisions of Title II of the National Housing Act. The aggregate amount of real estate loans may not exceed the capital and surplus of the bank or 60 per cent of the amount of its time and savings deposits, whichever is the greater.

The relaxation of restrictions on the making of real estate loans was a reactionary act. Such loans are lacking in liquidity and are in consequence a type of investment which the authorities have generally considered the commercial banks should avoid. The wholesale bank suspensions and other banking difficulties in the United States in recent years were aggravated by the unwarranted extent to which many banks had diverted funds to loans on real estate. However if at the time the loan is made adequate consideration is given, as encouraged by the requirements of the 1935 act, to amortization and other forms of security and to the long-time trends of the locality, the real estate loan can be a safe and desirable part of a bank's portfolio.

Speculative use of credit. Still another group of provisions in the act of 1933 attempts to set limits to the speculative use of credit. The requirement that federal reserve banks shall discount for members without discrimination is changed to say that they may discount for members; the reserve banks are hereafter to watch the use of credit by members, and the Board of Governors may put

an embargo on credit to any member found making an undue use of credit for speculative loans. Limitations on the making of loans on security collateral include a prohibition of the practice of placing such loans with brokers or dealers in the capacity of agent for a nonbanking organization. Leading bankers have remained skeptical of the effectiveness of these provisions to accomplish adequately their intended purpose. The stock market boom of 1936-37, with the precipitate crash of 1937-38, seems to support this skepticism.

Government control of credit. By the early 1930's, there were those in authority who believed that the defects of the American economy had gone so far that the maintenance of the then existing banking system might neutralize efforts to obtain economic stability, and that it was therefore essential to centralize control of the banking and credit system in the hands of an agency that could be held responsible for monetary policy. The appropriate agency, of course, was the Board of Governors of the Federal Reserve System. By means of the amendments to the Federal Reserve Act in 1933 and the revision of the banking laws in 1935, complete control over national credit policies was bestowed upon the Board of Governors, subject to Treasury and general government policies.

This legislation gives practical expression to the change that has taken place in the accepted concept of credit control in the United States. The founders of the Federal Reserve sought, by a strict definition of paper eligible for rediscount, to maintain a qualitative control of credit. The present function of the federal reserve banks is no longer merely the accommodation of trade by means of rediscounting short-term commercial paper. The system is now devoted to promoting business stability, which means that at times the Board of Governors, as a means of curtailing a boom or slump, may pursue a credit policy in direct contradiction to the immediate needs of business. The regulations permitting the substitution of government securities for commercial paper as a basis for the issue of federal reserve notes are consistent with this over-all policy.

Aside from the changes in the basis of note issue, the new policy has been fortified by permitting federal reserve banks to make advances to member banks on paper ineligible for rediscount, by centralizing control of open-market operations in the Federal Open-Market Committee stationed in Washington, and by giving the Board of Governors the right to change member bank reserve

requirements, discount rates, and margin requirements on security loans. The authority of the board has been further enhanced by its power to regulate the payment of interest on time and savings deposits and to exercise supervision over the use of credit for speculative purposes.

The present banking structure. The foregoing survey of recent banking legislation shows that some of the pre-existing weaknesses of the American banking system described in the immediately preceding chapter have been remedied but that at the same time certain new elements of weakness have been introduced. The table following shows the composition of the banking system as a whole as of June 30, 1947.

BANKING SYSTEM OF THE UNITED STATES[1]

DECEMBER 31, 1947

	Number	Deposits (in millions of dollars)	Loans and Investments (in millions of dollars)
Member Banks			
National Banks	5,005	82,023	65,280
State Banks	1,918	40,505	32,566
Total Member Banks	6,923	122,528	97,846
Nonmember Banks			
Mutual Savings Banks	530	17,747	18,614
Other Nonmember Banks	7,261	21,575	18,438
Total Nonmember Banks	7,791	39,322	37,052
Total All Banks	14,714	161,850	134,908

[1]Figures taken from *Federal Reserve Bulletin*, May, 1948, pp. 534–5.

While the total number of banks has been reduced from 25,330 in 1929 to 14,714 in 1947, there still remains that division of the system into many independent local banks of small average resources, which has always represented a fundamental weakness in the banking structure. The competitive chartering of banks by both federal and state governments is an American incongruity.

The problem has been diminishing in magnitude as a result of many schemes devised to strengthen the system of centralizing control. The number of branch banks has been steadily increasing, though their activity is still confined at best to the one state in which they are established.

Membership in the Federal Reserve System has been growing, slowly but definitely. Today 47 per cent of all banks are members of the system, while 76 per cent of total deposits are in member banks.

The establishment of a nationwide system of deposit insurance, to which practically all commercial banks subscribe, has extended the federal government's control over the banking system, and has tended to raise banking standards. Moreover, deposit insurance has helped to diminish the disastrous effects of bank failures both on the individual depositors and on the economy as a whole. With 12,071 banks on the Federal Reserve par list, check collections have been facilitated and another disadvantage of decentralized banking overcome.

Finally the increased centralization of control in the hands of the Board of Governors of the Federal Reserve System, as a result of the Banking Act of 1935, has removed the most serious objection to a regional banking system in comparison with the single central bank system that prevails in most countries.

The past two decades have witnessed striking shifts in the operations of the commercial banks. Bank profits, which held at low levels during the 1930's, rose during the Second World War to the level of the 1920's for the first time since 1929. Excess reserves, having reached unprecedented heights during the 1930's, were reduced during the war period to insignificant proportions. Although total earning assets in 1947 were about three times larger than in 1929, banks profits, owing to the fall in interest rates, were about the same in the two years.

While during the 1920's, earning assets consisted primarily of high-yield loans, they are now composed predominantly of low-yield government securities. Commercial banks are thus no longer principally engaged in financing industry by means of short-term commercial loans but are devoting most of their efforts to handling the financial business of the federal government. After the end of hostilities in 1945, there developed a moderate reversion to business financing as the Treasury began to retire the public debt and commercial loans increased to finance the reconversion of industry to peacetime production.

The federal government has in recent years played a larger part in the banking of the country. Not only has it entered directly into certain banking fields through the Reconstruction Finance Corporation and the agricultural and housing credit agencies, but the Board of Governors of the Federal Reserve System has turned from the passive policy of accommodating business to the active policy of seeking to control business conditions.

CHAPTER XVIII

Money and Prices

The value of money. As we have defined and used the terms "value" and "price," the value of anything is the quantity of some other thing that would be given in exchange for it, and the price of anything is the quantity of money that would be given in exchange for one unit of it. In practice it is customary to express values as well as prices in terms of money. But what then is the value of money itself?

The value of any commodity may be expressed in terms of any other commodity, but this is clumsy and confusing, and by common consent we have come to express all values in terms of money. In the same way the value of money might be expressed in terms of any other commodity. This would likewise be clumsy and unprofitable, but there is not here the same escape. We cannot obviously express the value of money in terms of itself. Moreover the concept of the value of money, to be useful, must not be its relation to some one commodity. What we are interested in is the relation of money to all other commodities and services, to economic goods in general. When we say money has depreciated in value, we mean that a unit of it will not buy so much. And we do not have in mind any particular commodity; we think of it as less valuable with respect to anything and everything that could be bought. So when money is of high value we mean that it will buy much, or conversely that much of other things must be offered to obtain any given quantity of money. This is evidently nothing else than what people mean by the "purchasing power" of money.

We thus arrive at a satisfactory working definition of the value of money and one that agrees closely with the popular concept. *The value of money is the quantity of other things in general that will be exchanged for one unit of money; i.e., its purchasing power.*[1]

[1] It will be noted by the careful reader that, as we have defined the term, value may relate to any quantity of the good in question (not merely to a unit). When we speak of the value of anything, as of land, we must specify the quan-

Value of money not constant. The universal custom of measuring all values in money, of noting how the prices of things rise and fall as expressed in money, is responsible for the common and persistent notion that the value of the monetary unit is a fixed thing and that all price changes are due to changes in the values of the priced commodities. A moment's consideration of the definition of money should convince anyone of the error of this notion. Value is always a relationship between two things. If one bushel of wheat was exchanged for two bushels of corn yesterday and for three bushels of corn today, we may say that the value of a bushel of wheat has *risen* from two bushels of corn to three bushels, assuming thus that the value of corn is a constant. But we could just as correctly assume the value of wheat constant and say that the value of a bushel of corn has *fallen* from one-half bushel of wheat to one-third bushel. Without knowledge of the facts of the particular case, the change in relation may have been due to a change in either corn or wheat alone or to changes in both of them. Under a money economy value and price are relations between money and other things, and it is always possible that a change in the value of anything may be due to a change in the value of money.

Measuring the value of money. But how is the value of money, or its purchasing power, to be measured? The answer must be found in that which expresses at once the relation of money to all other commodities and services. This is price, not the price of any particular thing, but a composite price of things in general, or what we have called the "general price level."

It is change in the general price level which indicates change in the value of money. A change in a particular price, while other prices remain substantially the same, is presumably due to changes affecting that particular commodity. But when the general price level rises or falls, we may safely assume that the cause is a change in the value of money. High prices and cheap money mean the same thing. Low prices and dear money are synonymous. A change in the general price level means an opposite change in the value of money.

tity, as a ten-acre farm, or a city lot. When referring specifically to the value of a unit of anything, we generally use the term "price." In referring to the value of a unit of money we cannot say price, since we have chosen to express prices in terms of money. Hence, unless otherwise stated, we shall use the phrase "value of money" to mean the value of one unit of money.

It follows that to measure and record the value of money we must have some way of expressing the general price level. For this purpose there have been developed *index numbers,* which are capable of reflecting with a high degree of accuracy changes in the general price level or the purchasing power of money. Before proceeding further in our study of the value of money and its relation to prices, we must acquire a working knowledge of the technique of index numbers.

Index numbers: Price relatives. We know that all prices do not move in unison. Some rise, while others fall, and still others remain stationary. A complete picture of all price changes in any given period would show a kaleidoscopic confusion of thousands of separate and apparently unrelated prices, rising or falling or standing still. Yet there is such a thing as a general price movement, just as there is a movement of a whole swarm of bees while the individual bees are dashing about in all directions in the utmost apparent confusion and without seeming relation to each other.

For a simple example, let us consider the average wholesale prices of three commodities, bituminous coal, wheat, and copper, in the three years 1913, 1914, and 1915. Here are the actual facts:

PRICES

	1913	*1914*	*1915*
Coal	$1.27 per ton	$1.17 per ton	$1.04 per ton
Wheat	.91 per bu.	1.04 per bu.	1.34 per bu.
Copper	.15 per lb.	.13 per lb.	.17 per lb.

Evidently the price of coal declined during this period, the price of wheat rose, and the price of copper fell from 1913 to 1914 and rose in the next year. To measure these separate price movements we may call one year, say 1913, the base year and express each price of each commodity as a percentage of the price of that commodity in the base year. For example, the prices of coal were: in 1913, $1.27; in 1914, $1.17; in 1915, $1.04. Calculating the percentages we have: $1.27/1.27 = 1$, or 100%; $1.17/1.27 = .92$, or 92%; $1.04/1.27 = .82$, or 82%. If then we let 100 stand for the price of coal in 1913, the price in 1914 relative to it was 92, and the 1915 price, 82. Such numbers are called *price relatives.* By similar calculations we find the price relatives for wheat and copper with respect to their prices in 1913, and the whole result is shown in the following table:

PRICE RELATIVES

	1913	*1914*	*1915*
Coal	100	92	82
Wheat	100	114	147
Copper	100	87	113

Such a table might be continued to include similar price relatives for all other important commodities, giving thus a record of the price changes of each separate commodity. But what we now desire is to know how prices in general have changed; we are seeking a record of changes in the general price level. For this there is required a number which shall express for each year the general or average price relative, evidently the average of the separate price relatives for each year.

The average, simple and weighted. Now there are many kinds of averages, and the particular average chosen will determine the character of the index number. For example we might employ the simple arithmetic mean of the separate price relatives of each year. In our example the average for 1913 is of course 100. For 1914 the simple arithmetic average is $(92 + 114 + 87)/3$, or 98. For 1915, it is 114.

Index numbers have been constructed on this principle, but they have a serious defect. The several commodities which enter into trade are not of equal importance, and an index number which takes no account of such differences is not an accurate record of the changes in the general price level or the purchasing power of money. The value of money is a good deal more affected, for instance, by a change in the price of wheat than by an equal change in the price of clover seed.

To avoid this defect of the simple arithmetic mean it is necessary to use a weighted average which takes proper account of the relative importance of the several commodities. For example suppose that in a given year the price of corn increased 4 per cent, while the price of butter increased 10 per cent, and that corn was twice as important as butter. The simple arithmetic average would show a combined increase of 7 per cent: $(4 + 10)/2 = 7$. But this is not a true picture, since no account is taken of the greater importance of corn. The average is corrected by putting the figure for corn in twice: thus $(4 + 4 + 10)/3 = 6$. This is a weighted average of the percentage increases.

Weighting may be accomplished in various ways, the most satisfactory method being to weight the price movement of each commodity according to the total value of that commodity exchanged in a given year. Thus if twice as much money is expended for commodity A as for commodity B in the selected year, this is taken to indicate that A is twice as important as B and should be given double the weight of B in the index number. The simplest way to accomplish this is to calculate first the total value of all the goods exchanged in a given year and then to compare with this the value of exactly the same quantities of these goods reckoned at the prices of each of the other years.

An index number of the aggregative type. We are now ready to apply this averaging technique to the prices in our previous example. Here are the facts for the first year:

YEAR 1913

	Prices in 1913	Quantities exchanged in 1913	Values at 1913 prices
Coal	$1.27	447,000,000 tons	$ 568,000,000
Wheat	.91	555,000,000 bu.	505,000,000
Copper	.15	812,000,000 lbs.	122,000,000
Total value			$1,195,000,000

It cost $1,195,000,000 to buy these quantities of these commodities in 1913. If the same quantities of these commodities cost more or less in 1914, it must have been because of changes in their prices, and if we could determine what the cost was in 1914 we should have a very good measure of the average movement of these prices. This may be readily calculated, as follows:

YEAR 1914

	Prices in 1914	Quantities exchanged in 1913	Values at 1914 prices
Coal	$1.17	447,000,000 tons	$ 523,000,000
Wheat	1.04	555,000,000 bu.	577,000,000
Copper	.13	812,000,000 lbs.	106,000,000
Total value			$1,206,000,000

Evidently commodities which cost $1,195,000,000 in 1913 were worth $1,206,000,000 in 1914. There must have been a small rise in the average price level of these three commodities.

For the year 1915 similar results are obtained, as follows:

YEAR 1915

	Prices in 1915	Quantities exchanged in 1913	Values at 1915 prices
Coal	$1.04	447,000,000 tons	$ 465,000,000
Wheat	1.34	555,000,000 bu.	744,000,000
Copper	.17	812,000,000 lbs.	138,000,000
Total value			$1,347,000,000

Again there has evidently been a rise in the average price level.

The final step in deriving a real index number of the prices of these three commodities is to express the total value of each year as a percentage of the total value in 1913. Thus: 1,195/1,195 = 1.00, or 100%; 1,206/1,195 = 1.01, or 101%; 1,347/1,195 = 1.13, or 113%. These percentages are the index numbers, and the series may be thus expressed:

Year	Index number of prices
1913	100
1914	101
1915	113

The formal definition of an index number of prices is as follows: *An index number of prices is a number which expresses the general price level for any given year or other period relative to the general price level of some particular period which is taken as the base period and whose general price level is commonly represented by 100.*[1]

The particular type of index number which we have here developed represents what is known as "an index number of the aggregative type." The year 1913 is the base period, its index number being 100. The weighting is according to the values exchanged in 1913. Though the same in our example, the year chosen for the weighting need not necessarily be the base year.

Practical index numbers. The example used in the foregoing explanation was purposely made simple by employing only three commodities. A practical index number must take account of a

[1] The term "general" in this definition does not imply that the prices of all commodities are included in any given index number. The meaning is that the index number represents the result of averaging price relatives for all commodities which have been considered. It is only in this sense that any index number shows a "general price level.' Actually it is possible to construct index numbers which are capable of serving almost as well as though they were in fact based upon all commodities.

large number of articles and may be calculated for any number of years, but no new principle is thereby involved. Also it should be noticed that other averages than the weighted arithmetic mean may be employed. Of the many possible forms of index numbers, the aggregative type both illustrates sufficiently the theory of index numbers and is probably the most useful of the types in actual practical use.

Two of the most useful series of index numbers and the ones most generally used in the United States are those published by the United States Bureau of Labor Statistics. One is the bureau's index number of wholesale prices and the other an index number of cost of living.

The wholesale price index was originally published in 1902, since which time it has undergone various changes in construction

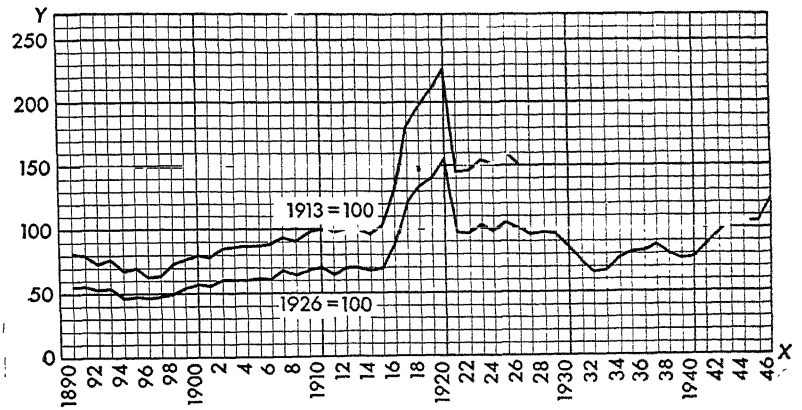


FIG. 29. INDEXES OF WHOLESALE PRICES

and commodity composition. Originally an unweighted mean of price relatives, it is now an aggregative index number. Beginning with 192 commodities for the first year for which it was computed, 1890, the number of commodities included increased to 297 in 1914, 404 in 1926, and 850 in 1946. The base year has been shifted several times and is now 1926, and the latest quantity weights employed are averages of amounts sold in 1929 and 1931.

The table on the next page presents the wholesale price index in two series, one on a base year of 1913, and the other based on 1926. It also shows, in the form of the reciprocals of the corresponding

WHOLESALE PRICES AND VALUE OF MONEY IN THE UNITED STATES (1890–1947)

Year	1913 BASE	1926 BASE	
	Index number	Index number	Purchasing power of the dollar
1910	101	70.4	142.0
1911	93	64.9	154.1
1912	99	69.1	144.7
1913	100	69.8	143.3
1914	98	68.1	146.8
1915	101	69.5	143.9
1916	127	85.5	117.0
1917	177	117.5	85.1
1918	194	131.3	76.2
1919	206	138.6	72.2
1920	226	154.4	64.8
1921	147	97.6	102.5
1922	149	96.7	103.4
1923	154	100.6	99.4
1924	150	98.1	101.9
1925	159	103.5	96.6
1926	151	100.0	100.0
1927		95.4	104.8
1928		97.7	102.4
1929		95.3	104.9
1930		86.4	115.7
1931		73.0	137.0
1932		64.8	154.3
1933		65.9	151.7
1934		74.9	133.5
1935		80.0	123.5
1936		80.8	124.2
1937		86.3	115.8
1938		78.6	127.2
1939		77.1	129.7
1940		78.6	127.2
1941		87.3	114.6
1942		98.9	101.2
1943		103.1	96.9
1944		104.0	96.2
1945		105.8	94.5
1946		121.1	82.6
1947		151.8	65.9

index numbers on a 1926 base, the purchasing power or value of the dollar in each year relative to its value in 1926.

For many purposes the most useful way to present an index number is by means of a graph or curve. The annual index numbers of wholesale prices furnished by the Bureau of Labor Statistics are thus shown in Figure 29 (page 359).

The second of the bureau's price indexes in general use is that of the cost of living. This index is also of the aggregative type and, as revised in 1940, uses the average prices of 1935 to 1939 as base. It includes 198 commodities in such groupings as food, clothing, shelter, and other items entering into family consumption, weighted according to actual family experience. To meet the extraordinary needs of a war economy, the index was revised during the war years. Although for certain purposes the cost of living index might be more useful, most monetary authorities continue to employ the wholesale price index to measure the value of money.

The problem of the price level. We are now ready to inquire into the forces that determine the value of money and the general price level. These forces operate through three factors; namely, the volume of trade, the quantity of currency, and the velocity of circulation of currency. These factors are all subject to more or less continuous and independent changes.

Volume of trade. The chief service performed by money is the making of exchanges. From the analogy of demand and supply and the value of commodities in general, it should be evident that the value of money is dependent in part upon the volume of exchanges, since it is this that measures the need of money. The more money work there is to be done, as measured by the volume of exchanges, the greater is the need of money, and the greater will be the value of a unit of it; and vice versa. As our first step we conclude that the value of money tends to vary directly with the amount of trade.

Quantity of currency. It is by now a commonplace that the value of a unit of anything (*i.e.,* its price) tends generally to vary inversely with the quantity available for use. In a season when there is an unusually large wheat crop we expect, other things being equal, to see the price of wheat comparatively low. A crop failure generally leads to a high price of wheat. So it is with money; the more dollars there are to do a given amount of money work, the less is the value of one dollar; and vice versa. In other words, the value of a unit of money tends to vary inversely with its quantity.

All of the money work, however, is not performed by money, and in determining the value of money, account must be taken of the other forms of currency. The only important one is bank deposits subject to check. The quantity of bank deposits has the same relation to the value of money as the quantity of money itself. The con-

clusion of the previous paragraph should therefore be modified to read: the value of a unit of money tends to vary inversely with the quantity of currency.

However the quantity of money and the quantity of bank deposits are not wholly independent magnitudes. As we have learned, the amount of deposits which the banks have outstanding is governed in part by the amount of their reserves. When the ratio of reserves to deposits is high, the banks are likely to encourage loans and so increase their deposits. When the reserve ratio is low, loans are curtailed and deposits checked. There tends thus to be at least a rough relation between reserves and deposits. And since the division of the total stock of money between bank reserves and other uses depends upon habits of the people which have a certain degree of persistence, we can discern a relation between the total quantity of money and the deposits of the banks. Moreover causes which would lead to increases or decreases in bank deposits, such as business activity or depression, tend also to have the same effect upon the volume of money, principally through changes in the amount of bank notes, and this contributes also to maintaining the relation between money and deposits.

While not neglecting this relation, we must avoid the opposite error of overlooking the possibility of considerable independent variations in the respective quantities of money and bank deposits. And we must therefore include both of these magnitudes in our analysis of the quantity of currency as a factor in determining the value of money.

Obviously the value of money can be affected only by such currency as is actually performing the monetary function of a medium of exchange. When, as sometimes occurs, the people are hoarding unusual quantities of money, such money is for the time being withdrawn from use and ceases to be a part of the quantity of currency that affects the value of money. Of course there is always a considerable amount of money in pocketbooks, in the homes of the people, and elsewhere, held for later use. The difference between quick spending and hoarding is after all one of degree. So long as the money so held is the normal amount as dictated by custom, such money should be considered as actually employed as part of the medium of exchange. Only an abnormal amount withdrawn for hoarding may properly be regarded as outside the sphere of influence upon the value of money. The variable influence of money normally withdrawn in varying degree from

active circulation is properly taken into account in connection with the velocity of circulation of money, to which we shall give attention in the next section.

In like manner, the value of money is affected by bank deposits only so far as they are available for making current purchases. This is why savings bank deposits and time deposits generally are not included.

Velocity of circulation. Unlike most other commodities, money is not consumed when used. A coin or note, after facilitating one exchange, is ready with unimpaired efficiency to perform another exchange.[1] The efficiency of a given quantity of money to make exchanges is thus materially affected by its rapidity of turnover. One dollar used or "turned over" twelve times in a given period comes to the same thing as twelve dollars used once apiece. The number of times on the average that a unit of money changes hands during a certain period of time, say a year, is called the "velocity of circulation" of money. The amount of money work that can be done in a given time is the product of the quantity of money multiplied by its velocity of circulation, and changes in velocity have the same effect on the value of money as proportional changes in its quantity. This principle is of course equally true of other forms of currency, in particular of bank deposits. It has been estimated that money changes hands on the average in the United States about twenty-five times a year. The velocity of circulation of bank deposits is estimated to be somewhere between fifty and sixty times a year.

The quantity theory of money: Algebraic relations. Summarizing these separate conclusions, we find that the value of money tends to vary inversely with the quantity of currency and its velocity of circulation and directly with the volume of trade. Since it is the general price level which measures the value of money, varying inversely with it, our conclusion may equally well be stated thus: *The general level of prices tends to vary directly with the quantity of currency and its velocity of circulation and inversely with the volume of trade.* This principle has come to be known as the "quantity theory of money."

The reader may possibly be helped to a clearer understanding of these relations by putting the several quantities into an algebraic equation. Let M stand for the average quantity of money in circu-

[1] The fact that paper money wears out is not of significance, since the worn-out notes are always automatically replaced by new ones.

lation during the year, and let M' denote the average amount of bank deposits subject to check and all other forms of currency. (The other forms are relatively insignificant and may be disregarded.) Let V and V' denote the velocity of circulation per year of money and of other forms of currency respectively. Let T stand for the volume of trade; *i.e.,* the total number of units of goods and services exchanged for currency during the year. Let P represent the general price level; *i.e.,* the average price paid for all the units of goods and services so exchanged. We then have the following equation:

$$MV + M'V' = PT$$

The quantity of money multiplied by its velocity of circulation is evidently the total amount of money paid in making exchanges during the year. Likewise the quantity of deposits multiplied by their velocity of circulation gives the total amounts of payments by transfer of deposits.[1] The sum of these two products is the total amount of payments in all kinds of currency. The volume of trade multiplied by the average price is obviously the total value of all things exchanged for currency. This magnitude and the total amount of currency payments are necessarily the same, and the "equation of exchange," as it is called, is the algebraic expression of this relation.

It is P which measures the general level of prices and (inversely) the value of money, and the effect upon P of any of the other magnitudes is clearly shown by the equation. Changes in M or M' or V or V' tend to cause similar changes in P; changes in T tend to cause opposite changes in P.

Significance of the quantity theory. Few principles of economic science have been responsible for so much learned dispute and so much popular misunderstanding as the theory of just what causes determine the general level of prices and, more specifically, what the relation is between the quantity of money and the price level. The battle has raged most fiercely about the so-called quantity theory of money and the equation of exchange. While anything like exhaustive study of this problem lies necessarily outside the limits of this elementary book, we can with profit undertake some further analysis of the quantity theory and the equation of ex-

[1] Strictly speaking, $M'V'$ includes also any other forms of currency that may have been used and which, for simplicity's sake and because of their practical insignificance in the problem, we shall hereafter ignore.

change, with the purpose of seeing clearly just what the principle means and some at least of the things that it does not mean.

The equation of exchange sets forth the necessary balance, over a year or some other limited period of time, between six specified factors. The quantity theory of money expresses the influence upon the price level of each of the other factors as a "tendency;" *i.e.*, as the change in P which would follow a given change in any one of the other factors, "other things being equal." But other things may not—very generally will not—be equal. For example, a decrease in the quantity of money (M) would decrease the magnitude on the left side of the equation; this would require an equal decrease on the right side, necessitating a decrease in the price level (P)—always assuming no change in any other factor. But a decrease in M might be counteracted by an increase in M', necessitating no change in P, or even leading to an increase in P. Changes in M might likewise be counteracted by opposite changes in V or V''. Again a decrease in the quantity of money might fail to cause the price level to fall because of a simultaneous decline in the volume of trade (T), which would make the necessary adjustment on the other side of the equation. Many other combinations of changes are possible. Failure to understand this mutual balance of the factors involved leads to many erroneous assumptions, the most common and pernicious being the naïve notion that any change in the quantity of money necessarily causes a similar or even proportional change in the price level. While this is in truth the tendency, history records many instances where an increase in the quantity of money has not increased the price level or has not increased it proportionally, and vice versa.

Again for the sake of clear thinking, it must be noted that the several factors whose influence upon the price level concerns us are not entirely independent of each other. Changes in these factors may, in part, work themselves out on each other rather than on P. We have already noted the relation which may exist between money and bank deposits (M and M'); so far as this relation is concerned the consequence is to strengthen the influence of changes in M or M' on P. An opposite consequence follows from the fact that a sudden great increase in the quantity of money might have an effect on velocity of circulation, money passing on the average more slowly because of the greater quantity in people's possession. A sudden great decrease might have the opposite effect, causing the limited quantity of money to be turned over more rapidly. As re-

gards the relation between M and M' and T, it is true that a sudden great increase in M or M' would very probably increase temporarily the purchases of the people and so increase T. A corresponding decrease in M or M' might decrease T. These relationships, tending to weaken the effect of changes in M and M' upon P, would generally be temporary in their effects, ultimately yielding to the habits of the people and other more fundamental influences.

Even more important is the relation between the volume of trade and the quantity of currency. There can be no question that an increase in the quantity of goods exchanged is very likely to have a material effect in increasing the amount of money and still more the amount of deposits, while a decrease in trade has the reverse effects. As we have seen, an elastic currency responds to business needs, especially through changes in the amount of bank notes (which are money) and bank deposits. Changes in T thus may cause similar changes in M and M', and such changes will offset the tendency of the changes in T to cause opposite changes in P.

Finally we must note the fact—although limitations of space prevent its demonstration—that the equation of exchange is chiefly useful in portraying certain long-run relationships. It is less useful in explaining short-run changes. This is in part due to the fact that the equation combines all transactions in one total (T) and averages all prices so as to give one price level (P), whereas it is the various relationships between separate prices, rather than their average as shown by the general price level, which are of chief significance as causes of the short-run disturbances which are constantly appearing in the economic system.

As thus elaborated and qualified, the quantity theory of money is generally accepted as a true statement of the long-run relationships between the quantity of currency and the general price level.

Practical limitations of the quantity theory. The practical utility of the quantity theory of money, either as a means of predicting price changes or as a guide to official control of the price level, is by no means so great as is sometimes assumed. This limitation is due both to the interdependence of the several factors in the equation of exchange, to which our attention has just been directed, and also to the practical difficulty of discovering the exact magnitudes of the various factors. With the exception of the quantity of money and the general price level, these magnitudes must generally be supplied by means of estimates and calculations which are subject to a wide margin of error.

Far-reaching effects of monetary fluctuations. Fluctuations in the value of money are fraught with the most serious consequences to all the people, inflicting fortuitous loss upon some and bestowing undeserved gain upon others. Rights to future payments are normally expressed in terms of the country's monetary unit. When the time for payment comes the debtor will pay and the creditor will receive money whose value may be quite different from that which existed when the agreement or contract was made. A farmer borrowed $10,000 at a time when wheat was worth $1.00 a bushel. If before the time for payment arrives there has been such a fall in the value of money that prices, including the price of wheat, have generally doubled, the farmer will be able to pay his debt in dollars worth only half as much as those he borrowed. When he made the loan, it required the sale of 10,000 bushels of wheat to produce $10,000. Now he need sell only 5,000 bushels to obtain the $10,000 needed to pay his debt.

All investments which consist of rights to receive stated sums of money, such as bonds, mortgages, annuities, insurance, etc., are subject to the same principle. The investor is compelled in spite of himself to gamble on the value of the dollar. If it goes up, the dollars he receives will buy more; if it falls, his dollars will buy less. Such investors were fortunate gainers from this cause during the long period of falling prices that ended in 1897. After that the most conservative investors saw their resources dwindle from year to year as the index number of prices rose. On the other hand, a period of drastic price decline may injure investors through the failures and bankruptcies of many business concerns whose securities they hold.

Those whose incomes are in the form of salaries and wages are affected by fluctuations in the value of money in the same way as are creditors. They have agreed to render services in return for a stated number of dollars, which as they are received may be worth more or less than when the agreements were made. Only as agreements expire and new ones are made can the parties undertake to make new terms conforming to the new monetary value, and these new agreements are equally subject to the vicissitudes of future changes.

It is thus that adjustment lags behind the fluctuations of the monetary standard on all agreements except those which are for immediate payment or for relatively short terms. Here there is the current adjustment, through raising and lowering prices, which produces those fluctuations in the general price level by which the

community adjusts itself as best it may to the changing monetary standard.

Monetary reform: Stabilizing the value of money. The principal objective of monetary reform is to establish and maintain stability of the value of money, which means of course stability of the general price level. The attainment of this objective is claimed to be both possible and desirable, since (1) such an objective is easily understood and would command popular support, (2) the criteria for its administration can be easily established, (3) stable relations would be assured between debtor and creditor and between recipients of fixed and changeable incomes, and (4) business would neither be encouraged nor discouraged.

Of the many practical difficulties encountered in achieving stable prices, one is that plans for this purpose are necessarily corrective or remedial rather than preventive. Changes in a wholesale price index number which are usually the signals for such corrective action as an expansion or contraction of credit do not appear until after the disturbances of which they are manifestations have made considerable advance.

Again any such program could not operate as a corrective of the multitudinous short-run and accidental variations in the wholesale price index. Certain of these variations are natural and normal adjustments among the various groupings of prices in our complex price system, while others may arise from limitations in the coverage or construction of the index itself. In any event it would be foolhardy for monetary authorities to endeavor to neutralize each and every such small change by the use of the somewhat clumsy instrument of general credit control.

There is moreover serious question whether an absolutely rigid price level would serve the needs of a dynamic economy. Perhaps what is needed is only a moderately stable level.

Is a stable price level—or a moderately stable price level—possible? How may such a price level be obtained? The limits of this book will not permit us to explore this problem much beyond the mere outlining of some of the more significant current suggestions for a stable monetary system.

"Managed money." During recent years it has been contended that the ideal monetary system should not rest upon standard money but should be a regulated irredeemable paper money system. The value of the monetary unit would be controlled by official action, for which the two most effective instruments are (1) con-

trol of the quantity of money and of bank deposits to maintain a uniform or otherwise controlled level of domestic prices, and (2) dealings in foreign exchange to maintain the value of the money in relation to the monetary unit of other countries.

Although the term "managed money" has frequently been applied to the recently adopted irredeemable paper money systems, not all these systems have been characterized by any important degree of management, while on the other hand there may be a good deal of management in connection with a gold standard system through control of discount rates, open-market operations, etc. There was for example far less management of the Continental bills of credit or the greenbacks than there was of the United States gold standard system in the twenty years before 1933. However, the leading nations which went off the gold standard in connection with the crisis of 1929 and the following depression have exercised a considerable control over the international values of their paper money currencies.

For example the British government undertakes to keep the value of the pound fairly stable in terms of dollars, francs, etc. This it accomplishes by purchases and sales of gold in the open market. These operations require a fund set aside for this particular purpose. Great Britain in 1932 established the Exchange Equalization Account, as its stabilization fund is technically known, for the purpose of stabilizing exchange rates and "managing the currency." The fund is controlled by a committee of the Treasury and the Bank of England. It is an immense fund, approximating 575 million pounds and consisting of gold, foreign exchange, and national currency. It is secretly managed. Similar machinery and stabilization funds have been established in the United States, France, Belgium, Holland, and Switzerland.

To take a specific example of the operation of such a stabilization fund, let us suppose that in England the value of the pound begins to decline in terms of the dollar. The British committee may halt the decline if it is able to buy pounds and sell dollars; in other words, if it offers to redeem pounds with dollars at some price. But if it is to redeem pounds in terms of dollars, the equalization account must either have accounts in the United States with private banks or with the federal reserve banks from which it can draw dollars, or else it must possess sufficient gold so that it may secure dollars from the United States Treasury by selling gold to the Treasury.

Though such management may succeed in maintaining stability of foreign exchange rates, and thus in stimulating international trade, it is doubtful if internal price stability in a large country can be maintained by such measures. At best this practice might be successful in small trading countries where the greater part of the production moves in international markets, and even then it would require the coöperation of the central bank authorities.

The "stabilized dollar." Some years ago Irving Fisher proposed his now well-known "commodity dollar" or "stabilized dollar." Adhering to the gold standard, this plan would make the monetary unit—the dollar—not a certain weight of gold as now, but a variable weight of gold, so adjusted from time to time as to keep its purchasing power virtually constant. Gold coin would not be used in actual circulation. The government would be, as now, the custodian of the country's monetary gold. There would be in circulation gold certificates and other forms of money reading in terms of dollars (not necessarily much different from the present forms), all of which would be redeemed in gold by the government on demand. The amount of gold to be given for a dollar would be officially determined from time to time on the basis of the index number of prices, so that a gold dollar would always purchase about the same quantity of goods in general.

This plan, especially if simultaneous adoption by most of the leading nations be assumed, might conceivably operate successfully under normal circumstances. Whenever wholesale prices rose, the weight of the gold dollar would be increased; when they declined the weight of the gold dollar would be decreased. Variations of individual prices with respect to each other, like the movements of the separate bees in a swarm, would in no wise be affected. But the movements of all prices in general would be stopped; the swarm as a whole would become stationary.

It is improbable that the plan would be strong enough to cope with drastic short-run price changes occurring in times of crises. However, if it succeeded in eliminating or even moderating the long-range fluctuations in the general price level, it would have a smoothing effect on the shorter fluctuations and make the strain of crises easier to withstand.

There are, however, certain more formidable obstacles to the success of the stabilized dollar plan. Perhaps the most serious is the political threat. It may well be questioned whether the official administrative board would long be free to execute the plan with-

out the interference of a popularly elected legislature. Moreover, as with most other plans, corrective measures could be taken under the plan only after the price change had occurred, so that their influence would be remedial rather than preventive.

The "commodity reserve" scheme. The so-called commodity reserve scheme has been offered by Benjamin Graham, Frank D. Graham, and others as a solution to the stable money problem. Under this plan dollars would be made interchangeable with a composite unit of staple, storable, primary commodities. Holders of dollars would convert them at a fixed rate into warehouse receipts, thus obtaining ownership of these units or composite commodities, and vice versa. It is claimed that the system would give assurance that the prices of the commodities being used as the monetary standard would always be stable.

The plan would function in the following manner: when the prices of the commodities in the unit fell, merchants would purchase these products, place them in warehouses, and exchange their warehouse receipts with the monetary authorities for dollars. The increased purchases would tend to bolster prices, while the increased money in circulation would also exert an upward influence on prices. Conversely when prices rose, warehouse receipts would be redeemed, the supply of money subsequently reduced, and the stored goods sold.

This plan likewise is open to criticism. If it is to be successful, the commodity unit should be composed of all important goods that enter the market, such as manufactured goods, and not only staple raw materials. The general public, lacking the required commodities, would be unable to increase its cash holdings in times of falling prices, so that the plan is really applicable only to dealers in the products concerned. Lastly the quantity of stored goods might become excessive and burdensome during a lengthy period of declining prices, while in an inflationary era the stored articles might not be available in sufficient quantities to forestall the upward price movement.

One hundred per cent money. Some authorities contend that the inability of the Board of Governors to control the credit supply is because of the present system of fractional reserves which enables the banks voluntarily to vary the volume of deposit currency. Therefore they advocate that banks be required to hold a full 100 per cent reserve against all their demand deposits, a plan known as "100 per cent money."

Under this system additions to the supply of currency, new demand deposits, would be made only when new reserves were created by an increase in voluntary savings deposited with the banks. The volume of reserves would be controlled by the open-market operations of the Federal Reserve as at present.

The successful application of this plan entails the abolition of the short-term loan market now provided by banks of discount. It is generally believed that control over credit expansion and contraction is just as likely to be achieved by less drastic methods without handicapping businessmen by the prohibition of short-term lending.

Status of monetary reform. For the immediate future at least most of the nations of the world seem committed to managed money, involving a paper issue not redeemable in standard money on demand, regulated foreign exchanges, and internal credit control. In support of this system, there is marshalled a considerable weight of theoretical argument based upon the quantity theory of money, but there is theoretical argument also against the plan.

In particular, the hazards of a monetary standard controlled by government and so always a potential football for political contest are serious. The temptation to pay the costs of government by paper money issues and bank credit, in lieu of the less popular taxation, and the popular sentiment in favor of cheap money, which is so prevalent in many countries, furnish two powerful forces which as a matter of fact have eventually directed most irredeemable money systems down the path of uncontrolled inflation.

It may be observed that these dangers are especially to be feared in the United States. Historically the world's experience with irredeemable money has been, almost without exception, disastrous. It is true that the United States has now maintained an irredeemable paper money system, involving a certain degree of "management" for at least a decade without bringing disaster. At the time of this writing, the prospects of the managed money systems of other nations are anything but bright.

Monetary theory is not wholly clear even as to what should be the aim of a managed money system. The argument for a constant price level is not so conclusive as may at first appear, since in our existing economic order a stable price level might conceal the presence of factors making for eventual disequilibrium. Neither can it be said that the value of the monetary unit ought to be maintained at any specific level in terms of gold. Since Great Britain

went off the gold standard in 1931, the gold value of the pound has in fact fluctuated.

It is proposed by some that the managed money system be employed, not to preserve a stable price level, but to regulate the price level so as to bring about desired ends embraced in the government's program. This takes the problem into a realm in which economic science does not yet speak with sureness. In spite of the recognized defects of the gold standard and the attractive features of some of the current monetary proposals, as a matter of practical policy the presumption is still in favor of the gold standard.

EXERCISES

1. Compute an index number of prices for each of the three years, on the 1926 base, using the simple average of price relatives:

	1926	1932	1938
Wheat (bu.)	1.20	.40	.60
Cotton (lb.)	.15	.06	.09
Coal (ton)	5.00	4.00	7.00
Iron (ton)	12.00	9.00	15.00
Petroleum (bbl.)	1.20	.60	1.00

2. What is the change in the value of money if prices rise 50 per cent? If prices rise 25 per cent? If prices fall 20 per cent? If prices fall 50 per cent?
3. In terms of the "equation of exchange" explain what effect each of the following would tend to have upon the general level of prices in the United States:
 (a) The issue of a large amount of paper currency by the federal government.
 (b) Increased productive efficiency of American industry.
 (c) Open-market purchases of government securities by the federal reserve banks, increasing the size of commercial bank reserves.
 (d) A decrease in borrowing by business men.
 (e) Improved facilities for transfer and collection of checks.
 (f) Expectation of rising prices.
 (g) A law increasing the size of the reserve ratios which banks must maintain.
 (h) Floods, droughts, strikes.
 (i) Increased volume of loans by banks to businessmen.
 (j) A sudden wave of thrift, or an increased desire to save.
4. *(a)* In the last stages of inflation in Germany after the First World War, prices rose more rapidly than the increase in the quantity of money. Account for this situation.
 (b) In the three-year period, 1930–32, the quantity of hand-to-hand currency in circulation increased by more than 800 million dollars, yet prices declined severely. How, if at all, can this result be reconciled with the quantity theory of money?

CHAPTER XIX

Demand for Productive Factors

Functional distribution: Meaning. The purpose of economic activity is to produce income. The ultimate outcome of organized economic activity in any social system is to distribute the income thus produced among the members of the social group. This process of dividing into individual shares the total social income produced by coöperative effort is known in economics as "distribution."

A study of distribution is concerned primarily with *types* of income resulting from the productive process, rather than with the actual amounts received by specific individuals. There are four of these types: economic rent, interest, wages, and profits. For each type there is a corresponding factor of production to which that block of income is assigned: economic rent accrues to the landowners within society as a remuneration for the use of their land in the productive process; the total interest income goes to those who provide capital; wages go to the workers of all sorts—managerial, skilled, unskilled; profits to the entrepreneurs.

The productive factors and their incomes. A few words of definition will serve to clarify these concepts of types of income and their corresponding productive factors. First, we must define land and economic rent. The term "land" embraces all the resources of nature afforded by man's physical environment, including the mineral deposits of the earth, the forests, the water power, the medicinal springs, the fisheries, etc.

The term "economic rent" is a scientific term not to be confused with the word "rent" as used in daily speech. An ordinary lease or rental contract usually covers articles of wealth—in a city, buildings, sidewalks, and other improvements; in rural districts, fences, barns, dwellings, orchards—which are the result of man's labor and not a gift of nature.[1] The single rental paid for this composite of land and man-made instruments of production is composed of two

[1] Under certain conditions existing articles of capital may respond to the forces of demand and supply, over short-run periods, in much the same way as

different elements: interest on capital, and economic rent. The definition of economic rent is as follows: *Economic rent is a payment made for the use of land, distinct from all payments for labor and the use of capital.*

Capital consists of man-made instruments of production in all their many forms: ships, factories, transportation systems, vehicles, machines, tools. Their essential nature and the conditions under which they come into existence were discussed in Chapter II of this book. The process of capital accumulation begins when people decide to set aside a part of their current incomes for investment as opposed to consumption. These savings provide the funds available in the loan market and are transferred to borrowers on terms which provide for their eventual repayment with an additional premium known as interest.

From the standpoint of the entrepreneur, the acquisition of capital (except for his own capital devoted to the business) has its origin in a loan which he obtains on these terms. Since we propose to approach the problem of distribution from the businessman's standpoint, we shall focus attention on the loan market and, for purposes of the present analysis, consider loans and capital as synonymous terms. From this view of the problem, *interest is a payment made for the use of loan funds.*

Labor in the broadest sense of the term includes all exertion put forth as a means of securing income; hence there are few people who do not perform labor of one sort or another. Even in its narrower sense of labor performed by workers in the employ of other people, the term embraces a great variety of services ranging from those of the high-salaried business executive to those of the common unskilled worker. We shall use the word "wages" to define all payments for the services of workers, regardless of the type of service or the standing of the individual in the hierarchy of employment.

The entrepreneur, whose place in the modern industrial system was also described in Chapter II, performs the function of devising, organizing, and directing the productive unit, bringing together for this purpose the other productive factors, land, labor, and capital. He is usually himself a laborer in the enterprise, hence entitled to wages; he may also advance his own funds and provide his own

land. When this is true, the incomes from these capital instruments are determined by the principles of economic rent rather than those of interest. Attention will be given to this phase of the problem at a later stage.

land, in which case he is entitled to interest and rent. But if he receives anything in his capacity as entrepreneur per se, it is an income apart from these other receipts, consisting of *profits—a residual income accruing to the entrepreneur after all other charges against the revenues of an enterprise have been met.*

This is admittedly a simplified picture, assuming as it does that wages, rent, and interest always appear in connection with a transaction between individuals. There are numerous exceptions. A farmer, for example, who owns his land, provides his own capital equipment, and hires no labor is at the same time landlord, capitalist, and laborer, receiving a composite income in which the different types are indistinguishable.

We shall postpone consideration of cases such as this, where the types of income are *implicit* in the situation, and begin our inquiry with the simpler case, where rent, interest, and wages are *explicit* payments by the entrepreneur to the providers of productive factors. This is not only simpler but in the modern industrial system, it is the more typical case. Moreover it will be shown in due course that what happens here to determine the sizes of the various incomes—rent, interest, wages, and profits—exerts a governing influence on these incomes in their implicit forms.

Now it cannot be said that these four categories of income, when added together, exhaust the total social product, for there are other types of income than the four mentioned. Some people, for example, are given a share as a matter of governmental policy or private charity, through pensions to the aged and incapacitated, grants to relieve distress or poverty, and the like. Some contrive to get a share by trickery or theft. The peculiarity of these other forms of income is that they are determined not by the operations of the economic system, but either capriciously, as in the case of theft, or by moral and political considerations, as in the case of public grants to the incompetent or unfortunate members of society.

If the entire matter were left to the free play of economic forces, all incomes would fall into the four categories listed; no one would share save as a contributor of one or more of the productive factors. For our purpose, therefore, a study of distribution resolves itself into a study of the different factors of production and of the economic laws which control their remuneration.

Distribution occurs in the process of production. This approach to the problem makes it clear that distribution takes place within

each productive enterprise as a necessary consequence of its routine operations. The wage earner receives his share from day to day or week to week in exchange for the sale of his labor. The landlord's share is determined in the typical case by negotiations with an entrepreneur, with the outcome recorded in the terms of a rent contract. Loan contracts, governing the terms upon which the entrepreneur acquires funds owned by others, determine the size of the interest income. Profits, the share going to the entrepreneur himself, are a residual income, that portion, if any, of the total receipts of the undertaking which is left in his hands after payment of all business costs has been made.

Consider, then, a specific business enterprise, representative of the great industrial world of today, engaged, let us say, in the manufacture of shoes. The business is privately owned and operated for profit. The owner binds himself to pay out definite amounts in rent, wages, and interest, and he accepts what is left over as his profit. The only source from which all these payments can be drawn is the money receipts of the enterprise, resulting from the sale of its product.

In the course of the year many other payments will be made for the purchase of raw materials, light and power, transportation, and the like. However, a moment's reflection will disclose that these payments constitute the money incomes of other businesses, to be distributed by them in rent, wages, interest, and profits. From the standpoint of the problem of distribution, these portions of the gross receipts of this particular enterprise can be ignored, since their distribution will be cared for at some other point in the industrial system, in accordance with the same laws which operate here.

It will also be useful at this point to set aside inquiry into profits, since profits are a residual income, whose size is affected by, but does not itself directly affect, the sizes of the other shares. We are, therefore, in a position to state our problem in the form of a simple question. What forces determine the rates of wages, rent, and interest which this entrepreneur is obliged to pay?

Rates of interest, wages, and economic rent are market prices. The immediate answer to this question is equally simple. These rates are prices, determined by the same kind of market forces as those described and analyzed in our study of value theory. In a competitive system each specific rate results from the interplay of

demand and supply in a market, in this case the labor market, the loan market, and the market for land subject to lease. Our first step, as in previous studies of price and value, must be to understand the conditions of demand and supply as they apply to labor, to capital, and to land.

On the supply side, these markets show distinctive peculiarities which cause them to behave differently the one from the other. It is self-evident, for example, that the supply of labor—*i.e.*, the number of workers of a given sort seeking employment at a given time—is governed by quite different forces from those which determine the amount of loan funds seeking investment at that time, or the amount of land available for factory sites. Likewise the conditions of supply for capital and land are dissimilar.

On the demand side, however, these markets have much in common. We shall begin our inquiry by analyzing the demand for the factors of production, leaving to later chapters a comprehensive study of their respective supplies.

Business demand for factors of production: Motive. We are thinking at this point, it will be recalled, of the behavior of a single buyer of the services of labor, land, and capital; in other words, of an individual demand. This buyer will assume that the prices he must pay for these services are predetermined. His problem will be to decide how much labor to employ at the prevailing wage rate, how much capital to borrow at the prevailing interest rate, how much land to rent. Nevertheless his demand, thus determined, becomes an element in the total demand of the market, by which, in conjunction with supply, these prices are fixed. In other words, the amount of his demand, though resulting in the first instance from the price, becomes a factor in setting that price. This is an aspect of market relationships with which the student has become familiar through previous study of price determination.

If his business is to prove a success, the entrepreneur must make his decisions in accordance with sound business judgment. He must employ labor, borrow funds, rent land only because, and to the extent to which, it is profitable for him to do so. His demand for these factors of production is a derived demand, derived from his desire for output from his enterprise. In terms of dollars and cents, he is willing, and can afford, to pay the price required to obtain a unit of one of these productive factors—be it labor, land, or capital

—as long as that unit can be expected to contribute to the money receipts of the enterprise anything in excess of its cost.

No elaborate argument is needed to prove that land, labor, and capital are productive factors. Without them, the enterprise could not operate at all; there would be no product. Moreover the amount of revenue to the business will vary in some relation to the amounts of all three factors, or of any one of them, employed in the undertaking. It is this productivity of the factors which provides the motive for demand.

Diminishing marginal productivity. But although it is easy to see that our manufacturer must gain possession of some land, labor, and capital, and that he expects them to produce—in the form of business revenue—the money with which their costs in rent, wages, and interest are to be paid, these general considerations do not tell us how much of each it is worth his while to obtain at the prevailing rates. To prepare the way for the answer to this question, we must take note of a principle which governs the productivity of each of these factors, the principle of diminishing marginal productivity.

Assume, for example, that our entrepreneur has equipped his factory with land and capital and is bringing together a labor force. In the early phases of this process, the product of the enterprise per laborer may increase as increasing efficiency results from better utilization of capital equipment. But the point will eventually be reached at which the unit productivity of labor will begin to fall as the number of workers increases.

There are two reasons for this. In the first place, the physical product per unit of labor will decline as each man in the expanding force is less well supplied with tools, machines, and other instruments of production. In the second place, beyond certain limits increasing quantities of output must be sold at lower prices, so that the reduced physical increment will be translated into money income at reduced rates. As a consequence the employer cannot go on expanding his labor force, land and capital being constant, without noting a decline in the amount of business revenue attributable to one unit of labor.

Labor, it is true, is not all of the same sort but is of distinctively different kinds—managerial, skilled, and unskilled—which differ in function and production. However, no single type of labor can be increased in quantity, all other factors constant, without eventu-

ally showing diminished marginal productivity. Moreover when the employer has devised that combination of the different types of labor which results in the greatest efficiency for the enterprise, to continue to expand the size of this balanced labor organization—land and capital equipment constant—will eventually result in diminished money returns from successive increments of expenditure on wages and salaries.

The same general principle operates in the productive exploitation of capital. The entrepreneur acquires this capital, in the first instance, in the form of loan funds, which he utilizes for the purchase or construction of buildings, equipment, and other instruments of production. If his enterprise is supplied with only a small quantity of capital, it will take the form of the most essential instruments. As the quantity of capital increases, other factors remaining constant, it will be applied to uses which grow less and less effective.

Whatever the form of capital instrument to which the entrepreneur devotes his borrowed funds, he will find each one subject to the law of diminishing marginal productivity expressed in terms of money income. And as in the case of his complex labor force, when he has designed that combination of the different instruments which results in the greatest technological efficiency, he cannot expand this integrated organization as a whole—other factors constant—without experiencing reduced increments of revenue from a given unit of borrowed funds. Similar analysis applied to the utilization of land for productive purposes will show that this factor also is governed by the same consideration.

As a summary of the preceding discussion, this principle may be defined in the following general terms: *After a certain point has been reached in the utilization of increasing quantities of a productive factor, the coöperating factors remaining constant, the product will increase in amounts less than proportionate to the increase of the variable factor.*

Law of entrepreneur-demand for productive factors. Given predetermined market rates of wages, interest, and rent, this attribute of the productive factors sets a limit to the amount of labor, capital, and land, which a given entrepreneur will demand at a given time. In the case of each factor his demand curve will have the familiar downward slope characteristic of demand curves for goods in general. It is a curve, not of marginal utilities as in the case of

consumable goods, but of marginal productivities. It represents the entrepreneur's estimate of the unit contribution to business revenue to be expected from varying quantities of the factor in question, and indicates by its downward slope that this contribution per unit varies inversely with the quantity.

The entrepreneur will be impelled to acquire larger quantities of a given factor when its marginal productivity is in excess of the rate he must pay. His desire for additional quantities will disappear when this discrepancy has been removed by the operation of the law of diminishing marginal productivity.

Accordingly the law of individual demand for any productive factor may be stated as follows: *The individual entrepreneur will expand his purchase of any factor of production at the prevailing market rates until the expected marginal contribution of that factor to the money income of his enterprise is equal to the price paid for it.*

Before going further it may be well to point out that we are dealing with the problem in the abstract and hence express our conclusions in theoretical terms. The law of individual demand here stated rests on the same assumptions regarding the accuracy of the buyer's judgment and the smooth operation of competitive forces on both sides of the market which underlie the general laws of value as developed in previous chapters. Adjustment of these assumptions to the realities of business experience will be made as our study proceeds. It can, however, be said with confidence that this law of demand accurately reflects the fundamental and controlling forces.

Total entrepreneur demand for the productive factors. The reader is familiar with the process by which individual demands are combined in a competitive market to form the total demand which, in conjunction with supply, determines price. In the markets for the productive factors, individual entrepreneurs, each of whose demand conforms to the principle of diminishing marginal productivity, create a total or composite demand, whose general character can be summed up in the following statement: *In a given market at a given time the rates offered for any factor of production will vary inversely with the quantity of that factor seeking utilization.*

Non-business demand for factors of production: General. The preceding discussion has been concerned only with demands arising from the industrial world, where the money costs borne by the employers of labor, the borrowers of loan funds, and the renters

of land are expected to be recovered by subsequent sale of products. But in society as a whole, demands for labor, land, and capital do not originate solely within the productive system.

Labor, loan funds, and land are put to many uses other than the production of goods to be sold on the market. In many of these cases the demand does not have any direct relation to marginal productivity as we have defined this term, and the rates offered for varying quantities of the factor do not conform to the law of demand given above. Yet demands from these other sources are influential in determining rates of wages, interest, and rent, and must be brought into our account of the forces operating in the markets for the factors of production.

Demand for labor. The principal example of such demand for labor is the commonplace buying of labor's services for the immediate satisfaction of human wants. The hiring of domestic servants, of entertainers, and in many instances of professional men, such as doctors and dentists, is a case in point. Here it is the utility of the worker's services which elicits the demand; the relation of the rate to the amount demanded is governed by the principle of marginal utility with which the student is familiar.

But these non-business rates of wages, though resulting from different causes, are not unrelated to the rates which prevail in the business world. One reason is that the demands of the business world and these other demands are often in competition with each other for the same workers.

Domestic servants, for example, can usually find alternative employment in factories and shops. Consequently, there is a tendency for them to transfer from one type of employment to the other if the rewards—including such nonmonetary elements as security, independence, congeniality of surroundings—are sufficiently attractive. A musician may offer entertainment to private audiences in direct exchange for fees, or may find employment in the band of a night club; a cook may prepare meals for a family table or for a hotel; a doctor can choose between private practice and employment by an insurance company.

In a highly industrialized society the paramount forces on the demand side arise in the business world; the rates of wages established there virtually determine the wages paid in non-business pursuits for such types of labor as can transfer from one market to the other, although these other pursuits, by drawing off a portion of

the labor supply, do have some part to play in wage determination. To take account of all possible cases we may add that the wage rates of workers for whom there is no demand at all in industrial enterprise are equated, through the operation of demand and supply, to the marginal utilities of their services to consumer-buyers.

Demand for capital: Consumption loans. In the capital or loan market, demands for non-business purposes are so constant and at times so vast as to constitute a major factor in fixing the rate of interest. It is necessary, therefore, to give this aspect of the problem rather extended discussion, in order to avoid misleading over-simplification.

At the outset, we call attention again to the fact that all demands for loan funds in the business world have one essential characteristic in common; they all reflect the borrower's expectation to recoup his costs from money receipts produced by the capital thus acquired. There is no such expectation in the case of borrowers outside the productive system. Here the cost of the loan must be met from other revenues, and the individual's demand must be governed by some consideration other than the marginal productivity of capital.

These loan funds are destined to be consumed and to disappear, leaving the borrower worse off in terms of future income, after interest and repayment charges have been met, than he would have been without the loan. For this reason such borrowing is often called *spendthrift borrowing*, behavior typified by the wastrel whose purpose is to live beyond his means as long as he can. This term accurately describes the loan transactions of many individuals. But there is a vast demand for loans of essentially similar character (in that the borrowed funds are put to nonproductive uses), to which a different characterization applies. This whole category of loans, therefore, will be called consumption loans. What motives prompt a demand for such a loan, and what factors determine its amount at given rates of interest in individual cases?

Time preference. As was pointed out in Chapter XIII, every loan contract represents a transaction in which present and future income are exchanged against each other. The borrower augments present income by the extent of the loan, and he expects to reduce future income by the payments of interest and repayment of principal. The effect on the lender's income stream is the reverse of this:

his present income is reduced; his future income is increased. There is no difference in this essential respect between loans made to businessmen, to spendthrifts, to governments, or to any other kind of borrower. Viewing the transaction in its entirety, the borrower always exchanges future income for present income, while the lender gives up present income in exchange for future.

Transactions which, in this way, link the future to the present call into play the human trait of *time preference; i.e., the inclination to place a higher estimate on present goods than on identical future goods.* This aspect of human behavior also was dealt with in Chapter XIII, but further emphasis on it here will be of value. Time preference results from man's impatience for present enjoyment—an attribute of human nature which causes us to experience a sense of sacrifice when obliged to postpone enjoyment and, conversely, to feel a sense of gain when, having reconciled ourselves to waiting for a promised pleasure, we are able to abbreviate the period of waiting. Although the pleasure is the same in magnitude, it grows in desirability with its nearer approach and hence is more highly prized.

While it is generally true that people place a higher estimate upon present than upon future goods, the degree of this time preference is by no means uniform. Each individual's time preference depends upon a number of factors, residing in his personal character and in the particular external circumstances in which he is at the moment placed. We may well note a few of the considerations that affect individual time preference.

Some men are by nature so impatient and impulsive, or so undeveloped in foresight and self-control, that present desires are unrestrained by the thought of future needs. Other individuals, though more stable in disposition, lack ambition and have on this account little motive for thrift and abstinence. All such people may be described as having a very pronounced positive time preference.

Love and a sense of duty toward dependents, on the other hand, tend to lessen positive time preference and may even bring about, at least beyond a certain point, a negative time preference. Thus a man may stint himself to buy life insurance for the sake of his family.[1] It is obvious, however, that not every man responds to this call of affection and duty.

[1] A life insurance policy is not necessarily an indication of negative time preference, for even in life insurance interest is paid. It may be at a low rate, but it

A person's age and situation in life have their effect on his time preference. A young man before marriage generally has a pronounced positive time preference. After marriage this tendency may be very much modified. Never having saved before, the young man may begin to provide for the future with great concern. But he may also be so absorbed in meeting increasing demands on his income that provision for the future seems impossible. An older man, feeling life slip by, may for that very reason develop a very high degree of positive time preference. On the other hand, he may be so interested in the future welfare of his dependents as to manifest, beyond a certain limit, a lower or even negative time preference.

The size and timing of people's incomes have a very important influence on their time preference. Individuals with large and stable incomes tend to have lower time preference than those with small or irregular incomes. Poor people generally have very pronounced positive time preference. When present income merely provides a comfortless existence, future needs are often forgotten. The size of the anticipated income is very important. The man who expects an inheritance or a future return on an important investment has less reason to curtail present expenditure. On the other hand, he who foresees the end of his own earnings as a productive worker, or the lessening of his income as an owner of wealth, is likely to be impelled to save.

It is sometimes convenient to be able to express time preference as a rate. This matter is precisely analogous to the concept of rate of interest which was developed in Chapter XIII. Thus if a person at any time has such time preference that $100 of income today would be to him of the same value as $106 a year hence, we say that his rate of time preference is 6/100 or 6 per cent.

Borrowing for consumption as determined by time preference. Whenever an individual's rate of time preference is higher than the rate of interest, he will be impelled to borrow. There is gain to him when he borrows under these circumstances, since the premium he must pay to acquire present income—determined by the interest rate—is less than what he is willing to pay, as determined by his rate of time preference.

But borrowing for consumption has the effect of enriching the present and impoverishing the future. Each unit of borrowing re-

indicates nevertheless the general prevalence of some degree of positive time preference, even among people who take out insurance.

quires a progressively larger future sacrifice. In accordance with the general law of diminishing marginal utility, the marginal utility of the present gain decreases, while the marginal utility of the future income sacrificed increases. It follows that the individual's rate of time preference, based as it is on the comparison of these two marginal utilities, tends to decline as he borrows, until eventually it becomes equal to the rate of interest. Thereafter borrowing ceases to be advantageous. Hence the amount an individual is willing to borrow at any given time tends to vary inversely with the rate of interest. His demand curve therefore is of the same general shape as that of the borrower for production, and at equilibrium his marginal rate of time preference is equal to the interest rate he must pay.

Borrowing of governments. In the discussion immediately preceding, we have been concerned with the consumption borrowing of individuals—actual human beings. There was a time, before the rise of the modern industrial system, when most borrowing was of this sort. Today, such borrowing by individuals is a relatively insignificant factor in the loan market. In the meantime, however, there has developed an enormous demand for consumption loans from another source, namely, that of political bodies.

The borrowing of political bodies is generally similar to the borrowing of individuals for consumption, although some government loans are for purposes of production. In the typical case, the government decides, on grounds of social or political expediency, to embark on a line of action from which no future financial returns can be expected. The deficit resulting from this expenditure is then covered by the sale of government bonds or other evidences of indebtedness. There is also much short-term borrowing, in anticipation of receipts from taxes whose income is unevenly distributed throughout the year. The government, like the private individual who borrows to obtain satisfactions in excess of his present means of payment, is in such cases merely anticipating future revenues in order to finance large present expenditures.

But though these government loans are similar in nature to the consumption loans of private persons, their amount is not determined by the same forces which control the action of the individual with reference to the loan market. The amount of loan funds demanded by political bodies does not vary in any predictable way with the interest rate. A high rate of interest will sometimes cause

the postponement of some government enterprise for which borrowing would be necessary; it is more likely to affect the tenure of the loan; *i.e.*, whether it runs for a long or short term. In time of war, the statesmen in charge of the nation's destinies simply borrow all that is needed to bridge the gap between tax receipts and war expenditures, or all that the lenders can be induced to provide for this purpose, paying the lowest rate of interest practicable.

Demand for loans. There are times, as during a major war, when the necessitous borrowing of governments is the controlling force in the loan market. Normally, however, the demand of entrepreneurs so vastly outweighs all other types of borrowing that it governs the market, and borrowers for consumption, individuals as well as governmental bodies, must adjust their demands to interest rates determined primarily by the marginal productivity of capital in the business world. Strictly speaking, however, all types of demand must be added together to give an accurate measure of the forces operating at any given time on the demand side of the market.

Demand for land. A complete analysis of demand for the productive factors must provide a place for demands for the use of land, arising outside the business world. A survey of the actual uses to which land is put in any typical urban community shows that only a minor fraction of the total area is employed in such a way as to produce a money income to the user. There are, of course, the sites upon which stand the stores, hotels, factories, and other commercial enterprises of the city, the roadbeds of the railways; the plots of the truck gardeners in the outskirts. Each such enterprise represents a demand for land, which conforms to the principle of diminishing marginal productivity.

But there are as well the sites of schoolhouses and other public buildings, the city parks and playgrounds, the public streets—above all, the private residences with their lawns and gardens. In none of these latter cases is the land so utilized as to create a revenue out of which rent can be paid.

Generally speaking, these nonmercantile demands for land can be grouped into two different categories: those governed by the principle of marginal utility, and those whose amount is determined by social or political considerations.

Typical of the first group are the lands embraced in private residences—the sites of buildings, the lawns and flower gardens, etc.

Such lands yield their utilities directly to the consumer and are in demand for that reason. Strictly speaking, each individual demand will respond to the diminishing marginal utility of these benefits from land as the area increases in size; that is, the individual, after the area of land he has acquired has passed certain dimensions, is willing to increase this area only if the cost per unit decreases. His demand curve, therefore, has the same downward slope which characterizes individual demands for all forms of consumers' goods, and, what is more to our immediate purpose, it is of the same shape as the businessman's demand curve for land.

It is impossible to formulate with precision the forces which determine the demand for land in the second category of non-business uses; *i.e.*, the demands of political bodies, charitable and educational organizations, and the like. In these cases the decision to acquire land is based almost entirely on noneconomic motives. The city government decides that it is good policy to set aside certain areas of land for parks or playgrounds and requires the citizens, as taxpayers, to provide the necessary money. Or a charitable organization decides as a public service to erect a hospital and acquires the necessary land by voluntary gifts from philanthropic persons. So it is with the plots of ground used as sites for schools and colleges, governmental buildings, churches.

These lands are devoted to their present uses as a result of demands controlled by humanitarian, political, or moral motives. There is often no definite relation between the amount of these demands and the rent, or the sale price of the land. (The student is reminded that the sale price of land is derived from its rent by the process of discount, or capitalization.) A high rent—or sale price—may, it is true, prevent a project of this character from being carried out and thus remove that particular demand for land; what is more likely, the project will seek a location in a part of the city where rents (and sale prices) are not prohibitive.

Viewing the matter from the other extreme, it does not appear true from general observation that universally low rents elicit increased demands for land for these philanthropic and political purposes. With regard to their relation to rent we can only say that these demands, predetermined as to amount, come into the market from time to time to withdraw lands from other uses, and by thus affecting the supply available for these uses, exert an influence on market rates of rent.

There is one peculiarity of land which is not exhibited to anything like the same extent by the other productive factors: *i.e.,* its immobility. A given area of land must be utilized in the place where it is found, regardless of the relative rent it commands as compared with rents in other parts of the country. Capital, in its original form of loan funds, will tend to flow to that point in the productive system where, other things being equal, the highest interest rates are paid. Labor of a given type moves about with less ease than capital, but will nevertheless tend to shift its locality in response to higher wages.

The physical immobility of lands tends to break down their markets into local units. In some of these, for reasons peculiar to that locality, a non-business demand may exert the decisive influence on rent. It is probable, for example, that the demand of governmental agencies for land in the nation's capital is the principal factor in the determination of rents and land prices in Washington, D. C. Similarly, a summer colony of wealthy people may place such a high marginal utility on land in a favored resort as to set the rates of rent prevailing there. However for the nation's land area considered in its entirety, demands for productive uses prevail over all other demands in determining the rates of rent.

Laws of demand for the productive factors: Summary. Since with few exceptions the non-business demand for the factors of production is incidental and secondary to the business demand, we may formulate from the preceding discussion the following statements of principles which, under pure competition, express the primary and controlling relations existing in the demand for these factors:

An entrepreneur's demand schedule for each of the productive factors, in accordance with their diminishing marginal productivity, indicates that he will offer decreasing rates for increasing quantities of the factor.

Given the market rate for a factor of production, the individual entrepreneur will tend to acquire such a quantity of that factor that its marginal productivity in his enterprise is equal to the rate paid.

The total demand in the business world at a given time for any factor of production shows declining rates offered for increasing quantities of that factor.

Given total supplies of the factors of production, conditions of demand will tend to fix a market rate equal to the marginal productivity of the factor concerned; *i.e.,* the market rate of wages will

tend to equal the marginal productivity of labor; the rate of economic rent, the marginal productivity of land; the rate of interest, the marginal productivity of capital.

At the point of equilibrium in pure competition, the country's supply of any given productive factor will be so distributed among the different enterprises that its marginal productivity will be the same in all of them. This statement, be it noted, refers to each factor considered separately.

Functional distribution through the business world. We are now in position to take a comprehensive view of the process by which the distribution of the nation's income is worked out, through the routine operations of the productive system. Although the total *real income* of a people consists, of course, in the quantity of goods and services appropriated for human enjoyment, it will be convenient to think of the national income in terms of money, the form in which the incomes of most people originally accrue.

We know that the people regularly spend a part of their money incomes on footwear of all sorts, making their purchases continuously, day after day, throughout the year. These transactions have the effect of transferring to the manufacturers of boots and shoes a definite part of the country's total money income. Similar activities are going on continuously in all other markets for goods and services. Money paid out in taxes is merely transferred to political bodies by whom it is spent. Even savings are normally passed over to borrowers, who pass them on either to the purveyors of consumable goods or to those who provide capital instruments. In other words, as the people of the country make purchases or investments from day to day, they set in operation the process by which their money incomes are distributed through the various markets to become the incomes of others who in turn set the same process in motion on another cycle of distribution.

It is the segment of the national income expended on footwear whose distribution we propose to trace as an example of the whole process. This income flows into the hands of the manufacturer as receipts from the sales of his product, and almost all of it is promptly passed on to other persons. We shall ignore at this place any part remaining as profits and follow the manufacturer's business income as it is paid out in the operating expenses of the enterprise.

For our purposes, these expenses can be classified into two groups.

The first of these includes all the manufacturer's expenditures on rent, interest, and wages; that is, his payments to the owners of productive factors used in his own enterprise. As these payments are made, they bring to an end this particular cycle of distribution in so far as money is concerned.

The second group of expenditures includes all the other business expenses: payments for materials, transportation, light and power, and the like. Although extremely varied in kind, all these expenditures have one characteristic in common; they represent transfers of income from one manufacturer to other entrepreneurs, who act as additional intermediaries in the process of distribution. They will be passed on again either to owners of productive agents, payments which mark the final stage in the process of distribution, or to other businessmen, who take up the process of distribution at this point and carry it further.

As we follow these business activities through to their conclusion, we can picture the money income originally spent on footwear as a series of streams flowing through the industrial structure, portions being drained off at this point and that into the hands of the owners of productive factors, until the whole amount is finally distributed into rent, wages, and interest.

If in this brief statement we appear to have overlooked any part of the entrepreneur's disbursements—such as his tax payments—it is not because these items of expenditure constitute exceptions to the principles we have outlined. They merely introduce another intermediary into the process of distribution. These disbursements, too, are finally resolved into payments to the owners of the productive factors.

The total amount of income thus distributed is determined in the first instance by the total value of footwear sold on the market. The amount transferred by the footwear industry to other entrepreneurs is determined by the value of their products bought by the manufacturers of boots and shoes. And so on down the line; each transfer of portions of this income to other intermediaries in the process of distribution is a market transaction governed by the laws of price. Furthermore, the amounts paid out at each stage in final distribution to landowners, workers, and lenders are also determined by market prices, the rates of rent, interest, and wages.

In strict theory, these rates at all points in the productive system are equated to the marginal productivity of the factor—or subdi-

vision of the factor—in question. Thus we can say, on this theoretical basis, that each worker receives a share of this portion of the national income which is equal to the marginal contribution of his labor group to the income; and so, too, with each landowner and each provider of capital.

By stressing the word "theory" in these statements, we have in mind that competition is often imperfect and that conditions arise which disturb the nice adjustment of actual business experience to the postulates of economic law. These qualifying circumstances will be given attention in due course. The point to be stressed here, however, is that the basic laws of distribution are in constant operation, tending against all frictions and interferences to parcel out the national income among the productive factors in proportion to their respective marginal productivities.

The nature of profits. A final question remains: If our analysis is valid, why should there be any share of the income left over in the form of profits to the entrepreneur? If each of the other three factors receives a return equal to its marginal contribution to the income, haven't they taken all there is to be divided?

One plausible answer might be that the entrepreneurs constitute a fourth productive factor which gets a share based on its own marginal productivity. We must bear in mind, however, that the services of the entrepreneurs as active managers are paid for in (implicit) wages; if the entrepreneurs supply capital or land, they are paid in (implicit) interest and rent. When we speak of profit—pure profit—we refer to an income apart from, and in excess of, these other receipts.

Anticipating a fuller discussion of the question in a later chapter, we may say here that profits (and losses) result primarily from the uncertainties of business—the risk element. In a riskless economy, with perfect competition working continuously, there would in theory be no such income as pure profit. In such case wages, interest, and rent would absorb the entire income.

If pure profit exists as a type of social income—and there is no reason to question its existence—the explanation must be sought in conditions which fall short of this theoretical assumption; particularly in the dynamic character of the economy, the imperfections of competition, and the uncertainty of future events, which make business risky.

CHAPTER XX

Economic Rent

Introduction. In the preceding chapter the productive factors and the types of income which accrue to them through the processes of distribution were identified. The considerations which give rise to entrepreneur demand for the productive factors were discussed, as were the forces which, operating on the demand side of the market, define the prices or rates of return which enterpreneurs are impelled to offer for varying quantites of these factors. We turn now to a closer examination of land, capital, and labor and the characteristics peculiar to each which determine their respective shares of the social income—economic rent, interest, wages. We shall begin with land and economic rent, the income of landowners.

The productivity of land. When the subject of economic rent was introduced in the preceding chapter, it was assumed that land is productive—*i.e.*, capable of making its own specific contribution to the product of industry—and that the entrepreneur has an eye to this specific contribution when negotiating a rental contract. This assumption requires further demonstration than it has as yet received.

Approaching the matter from the point of view of society as a whole, the question may be put in this form: Do the land resources of a country, when exploited, make their own specific contribution to the social income distinct from the contribution made by the other productive factors? If so, can this contribution be measured in precise terms for the individual units of land?

Taking the general view of the problem, the answer can be found by assuming that a unit of land which is being employed for productive purposes is withdrawn from use, and that the labor and capital formerly applied to its exploitation are transferred to other uses. If the social income remained undiminished, it could be said that this land has contributed nothing; and on this assumption its owner could not induce anyone to pay him an income for the privilege of using it.

If, on the other hand, the social income is reduced by the withdrawal of this land from use, the amount of the reduction may justly be attributable to the land, and not to labor, capital, or any other productive agent, since these other agents have remained unaltered in amount. Also in this case the landowner can demand that income be given to him in exchange for the use of his land. Other members of society can afford to pay him for the use of his land, and they will still be as well rewarded for their labor and capital as before. Moreover if no one is willing to contract for the use of his land, the owner can gain this same income—in addition to payment for labor and capital—by exploiting it himself.

In any well-developed economy most natural resources are being exploited beyond the point of diminishing returns. If part of the land is withdrawn from use, and the capital and labor thus released is used elsewhere, the total output of the productive system will decline. In their new location these productive agents, capital and labor, will operate at lower marginal productivities. The decline of product is obviously attributable to failure to exploit the land which, in our illustration, has been thrown out of use; and it is an accurate statement of the facts to say that the quantum of product thus lost was the specific contribution of the land in question. This being true, the owner of the land will, as stated above, be in a bargaining position which enables him to exact a payment for the use of his land, a payment equivalent to its specific contribution.

Agricultural lands: Diminishing returns and marginal cost. We have now to see how the specific productivity of land is resolved into money receipts for the landowner through the normal processes of the market. We shall first consider this matter as it relates to agrarian lands, and to simplify the problem, we shall take the case of a landowner who is exploiting his own land and who, if he receives economic rent, must obtain it directly from the market in which he sells his product.

Let us assume that a man who owns a fifty-acre farm is raising oats for sale in a competitive market. The exploitation of this small farm will conform to the principle of diminishing returns; that is, successive applications of labor and capital to this fixed area of land will, at some point in the operation, return diminishing marginal products.

For purposes of illustration, we shall translate these successive

applications into money costs. The costs must include: compensation for the working proprietor equivalent to normal wages for men of his skill, normal interest on money invested in tools, machinery, and other capital instruments; also, all expenditures on running costs—hired labor, seed, fertilizer, and the like. We shall assume that all these costs of operation can be resolved into successive units of expenditure of $100 each.

Each such unit will produce a certain output in bushels of grain, but beyond the point of diminishing returns, this output per unit of expenditure will decline. The illustration can be stated in numerical terms as shown in the following table. We have only to state the matter in this form to see that diminishing returns from the use of lands is merely an alternative way of expressing the familiar concept of *increasing marginal cost.*

INCREASING MARGINAL COST IN AGRICULTURE

Expenditure in dollars	Total product in bushels	Marginal product in bushels	Marginal cost per bushel
$100	500	500	$.20
200	1,000	500	.20
300	1,400	400	.25
400	1,700	300	.33⅓
500	1,900	200	.50

The third and fourth columns in the table record the marginal product and the marginal costs of the farm under varying degrees of intensity of cultivation. The marginal product is the number of bushels attributable to a given unit of expenditure. For example, if three, rather than two, units of expenditure are applied, the total product will increase from 1,000 to 1,400 bushels; the increase—400 bushels—is the marginal product of a three-unit operation. Similarly, if four units are applied, the total product increases by 300 bushels, which is the marginal product of a four-unit operation.

The marginal cost is the cost per bushel of the marginal product. The marginal product of the three-unit operation is 400 bushels; this increment of product involved an expenditure of $100; the marginal cost, obtained by a simple process of division is 25 cents. The marginal cost of a four-unit operation—300 bushels obtained at a cost of $100—is 33⅓ cents; and so on.

The third and fourth columns of the table present alternative

ways of illustrating the principle of diminishing returns; first, in terms of declining marginal product; second, in terms of increasing marginal cost.

Marginal cost equated to market price. The bearing of this illustration on our main problem—economic rent—will become clear when we turn from a consideration of the costs of our farmer to take note of his receipts or income. Since he is producing goods for sale on the market, his money income will be determined by the price set upon his product through the operation of the forces of demand and supply.

Our farmer, for example, will plan to produce 1,900 bushels of oats only in the expectation of a future selling price of 50 cents per bushel; or 1,700 bushels only if the price is expected to be 33⅓ cents; or any other quantity only if the anticipated market price is high enough to cover the marginal cost of that quantity.

This statement in fact sums up the relationship of each individual producer to the market. Each assumes that the market price, at any given moment, is beyond his control and makes it his concern to avoid producing any part of his output at a loss. When he anticipates a rise in price, he is inclined to increase the application of labor and capital to the soil and thus advance his marginal cost toward equality with the new price, being induced to do so by the prospect of larger gains to be made from his increased output. Conversely, when he expects a decline in price, he tends to contract his production until he ceases to lose on the marginal units of his output.

Economic rent is producer's surplus. All units of the same commodity sell in a given market at the same price. Since this price is equated to marginal cost, it follows that, when production has passed the point of diminishing returns, the sale of the total product results in a surplus return.

The amount of this surplus, for any given quantity of product, may be discovered by comparing the money receipts from the sale of the product with the total cost of production. Assuming, for example, that the market price for oats is 33⅓ cents per bushel and that our farmer produces 1,700 bushels, he will receive a gross money income of $566.66. Reference to the table shows that his total cost is $400; hence he receives a surplus gain of $166.66.

The point should be emphasized that this part of the farmer's income is a true surplus over and above all payments required to

induce production of the quantity offered for sale. It will be recalled that the total cost of production includes interest on all capital used, wages for all hired labor, replacement of all materials consumed, and a remuneration for the labor and management of the farmer adequate to call forth his effort. To avoid the appearance of having begged the question by juggling the conditions of our illustration, we should point out that no reasonable assumptions regarding this normal reward of the farmer will change the results of our analysis. We may take any figure we please as a measure of this factor in the cost of production; it must still appear in our column of total costs.

If the farmer's product sold for no more than enough to cover his total cost (including the normal reward for himself), it would nevertheless prove sufficient to justify his operation of the farm. The fact that it sells for more than this is to be attributed to the law of diminishing returns, which causes marginal costs to rise, and to the business sense of producers, which impels them to make their operations as profitable as possible.

This surplus is economic rent. A synonymous term for it, and one which fits exactly into the conditions of our illustration, is *producer's surplus*. It is inevitably involved in the income derived from the cultivation of the soil after production has been pressed beyond the point where diminishing returns set in.

Explicit and implicit economic rent. In the preceding illustration the economic rent accrued to the farmer as an implicit income, a part of his general money receipts from the market. But it came to him in his character as an owner of land, not as a worker or manager or owner of capital instruments, for he received in addition a normal reward for these other functions.

Suppose now that our farmer wishes to retire from active labor and so offers his land for lease. He will of course expect his earnings as a worker to cease. We may assume that he turns his supply of capital instruments into cash and invests the proceeds in good bonds, thus perpetuating that part of his income which consisted of interest. In these circumstances his real loss of income upon retiring from the land will be equal to the producer's surplus or economic rent which he has been receiving.

This is what the land is worth to him as income-bearing wealth, and he will ask a contract rental of this amount from any prospective tenant. The tenant will be able to pay a rental of this amount,

for after subtracting this rental from his gross income he will have enough left to cover all costs of production and still retain for himself a normal reward for his labor and management. Under competition therefore contract rent will tend to equal the surplus gain to the owner implicit in his cultivation of the soil.

Lands of different quality. We have dealt thus far with a single plot of ground of uniform quality. If the earth's arable land were all of one quality, its cultivation would give rise to economic rent as soon as consumers' demand for the product had so stimulated cultivation that a part of the supply must be produced under conditions of rising cost. But lands, as we know, vary in productivity because of differences in natural fertility, in proximity to the market, in access to transportation facilities, and so on.

In business practice these differences in quality take the form of larger costs of production on some lands than on others. Whatever the crop in question, there are some lands whose combination of location and soil properties enables them to return the largest receipts for a given expenditure; others are somewhat less productive of income from the same expenditure; others still less productive. This is true of different parts of a single farm. Exhaustive analysis of the entire land area would permit its classification into many varying grades of productivity. The principle of diminishing returns will apply to the exploitation of each of these grades of land, and each will return an economic rent if certain conditions obtain with respect to the relation of marginal cost to market price.

A simple illustration will make these relationships clear. We will take lands of three grades devoted to the cultivation of oats, their relative quality being shown by differences in their initial marginal costs of production. Grade A is the farm used in our previous illustration. The inferiority of Grade B is shown by the higher marginal cost of the first units of output. Grade C is inferior to Grade B for the same reason.

GRADE A SOIL

Expenditure	Total product	Marginal product	Marginal cost
$100	500	500	$.20
200	1,000	500	.20
300	1,400	400	.25
400	1,700	300	.33⅓
500	1,900	200	.50

GRADE B SOIL

Expenditure	Total product	Marginal product	Marginal cost
$100	400	400	$.25
200	700	300	.33⅓
300	900	200	.50
400	1,000	100	1.00

GRADE C SOIL

Expenditure	Total product	Marginal product	Marginal cost
$100	300	300	$.33⅓
200	500	200	.50
300	600	100	1.00

It is obvious that unless the price of oats is higher than 20 cents a bushel only Grade A soil will be cultivated, since marginal costs on the other lands are above this price. At 20 cents a bushel there will be no rent even on the best land, since the quantity produced at this price would return no producer's surplus. Moreover it is clear that except for diminishing returns the price would never rise above 20 cents; inferior soils would never be cultivated; no economic rent would ever appear. These statements hold good because in the absence of diminishing returns unlimited quantities of oats could be produced at the lowest cost shown in the table—20 cents a bushel.

Let us assume now that the little society pictured here experiences an increase in the demand for oats because of an expansion of population. This greater demand, together with the appearance of diminishing returns when Grade A soils are called on for greater production, will eventually cause the price to rise—let us say to 25 cents. At this price cultivation will begin on Grade B soil, and economic rent will appear on Grade A; Grade C land will still be idle.

If the price rise continues until oats sell for 33⅓ cents, economic rent will appear on soil of the second quality, economic rent will rise on Grade A land, and the poorest of the three grades of soil will be drawn under cultivation. The price must rise to 50 cents a bushel before this third grade of land returns an economic rent.

Margin of cultivation. The foregoing analysis introduces a concept of importance in the doctrine of economic rent: the concept

of the *margin of cultivation,* which may be defined as *that stage of exploitation at which a unit of expenditure just pays for itself.* This last, or marginal, expenditure is applied under such conditions that it produces an income only equal to the cost. That this is true may be due to either of two causes. The expenditure may be the last of a series applied to a given plot of ground, its product being restricted by the law of diminishing returns, or it may be the first expenditure made on lands so lacking in fertility or so remote from the market that, at the existing price of the product, they are barely worth cultivating at all.

It is useful to distinguish these two cases by calling the one the *intensive,* and the other the *extensive,* margin of cultivation. The intensive margin of cultivation on any grade of land is reached when the product of the final unit of expenditure sells for no more than it costs to produce. The extensive margin appears on that grade of soil which it is just profitable to cultivate at a given time. Referring again to our illustration, with oats selling at 33⅓ cents, Grade C soil lies on the extensive margin; the intensive margin is represented by the second expenditure on Grade B and the fourth on Grade A.

Since marginal cost equals market price, there is no economic rent derived from the expenditures on either the extensive or the intensive margin. As was made clear in our preceding illustration, the position of both margins at any given time depends upon the selling price of the product. If anything occurs to raise this price, the intensive margin on all lands under cultivation will be raised, and the extensive margin will be pushed out to lands whose lack of productivity made them extra-marginal at the old price. A fall in the price will, on the other hand, cause a lowering of the intensive margin on all lands, and a narrowing of the extensive margin by the process of throwing the poorest lands out of use. It should be noted that these statements hold good only on the assumption that costs of production remain unchanged as the selling price fluctuates.

Alternative uses of agrarian land. It is obvious that we have simplified the problem by assuming that only one crop is produced on the land. But it is really unnecessary to consider all the complexities caused by the vast number of different crops actually under cultivation at a given time, for this situation does not alter the principles of rent. For each of these uses the different parts of

the land area vary in profitableness because of differences in fertility, in location, etc. Under given conditions of cost and price there will be for each piece of land one most profitable crop and an intensive margin measuring the largest amount of that crop which it pays to produce.

Each farmer will try to employ his lands so as to expand his producer's surplus to the maximum. He will plant that crop which at the time offers the largest producer's surplus and will continue this line of production until some other crop becomes more profitable. In this way the existing area of agrarian lands is allocated among the different lines of cultivation so as to produce the largest economic rent under prevailing conditions of cost and market price.

Economic rent of urban lands. We have attempted to simplify an essentially complicated problem by limiting our discussion to economic rent from agrarian lands. But the same type of income arises from the exploitation of urban lands; it is produced by the same forces which create agrarian rents; its amount in any specific case is also controlled by similar forces. We shall not attempt to deal with all the complexities of this problem. The student will have grasped its fundamental principles if he attains a clear understanding of a few propositions. First, productive exploitation of urban lands conforms to the law of diminishing returns. Second, urban lands, like agrarian, vary in inherent productivity for any use to which they may be put. Third, market prices of goods and services whose production involves the use of urban lands are determined by costs on the margins of utilization. Fourth, the owners of urban sites appropriate the economic rent which accrues from the exploitation of the land.

Certain general characteristics of urban lands should be noted. They pertain to lands employed in connection with urban enterprises—factories, banks, shops, hotels, and the like. As implied in this definition, the service rendered by urban lands is that of providing sites for buildings and other capital instruments. When we speak of the productivity of urban lands in a given use, the term does not refer primarily to the fertility of the soil, but to its location. For this reason economic rent of urban lands is denoted by the term *site rent*.

Each urban enterprise will find that the plots of ground within the city vary in effectiveness to serve its purpose. Some will provide

locations of maximum advantage; the location provided by others will be less advantageous; still others will be so badly situated as to be useless to the enterprise. The factors which determine the relative effectiveness of land sites are, of course, not the same for one enterprise as for another; retail stores, for example, require attributes of site for effective location different from those required by apartment hotels. But given adequate information as to the technical requirements of any enterprise, together with information regarding the general features of the city—density of population, layout of streets and traffic systems, and the like—it is possible to classify the land sites in regional groupings according to their relative effectiveness for that enterprise.

When a city of any considerable size has become settled, it is not difficult to see that the retail stores, the theatres and hotels, the wholesale establishments, the banks, and other enterprises have spread out over sites which vary in their effectiveness for the business in question. It will be evident that Merchant A has a better location than Merchant B. If A were obliged to move his store to B's site, his receipts would fall off, even if he used as much capital and labor as formerly and applied managerial ability of equal skill. There is no way to explain such differences in the profitableness of like enterprises differently located than to attribute them to the income-producing powers of the land.

The principle of diminishing returns operates on each plot of urban land whatever the use to which it is put. For example in retail trade an increased expenditure of capital and labor on a site of given dimensions takes the form of additional stories to the building, a larger area of floor space, an increased stock in trade, increased working force, a more complex and unwieldly organization. As the building grows in size, the costs of maintenance, of transportation of goods and passengers from floor to floor, of supervision, will eventually rise. Many small sources of waste, which in a smaller establishment would be discovered and corrected, will pass unnoticed by the managers. For many reasons, the costs of rendering the service will rise, or, to use the correlative expression, the returns from the last unit of expenditure will fall. So it is in other urban enterprises. As a result, the time must come when additional expenditure on the best sites for a given use no longer affords a larger profit than the same expenditure on lands not so well situated.

The exploitation of each parcel of urban land tends to be pushed to the limit, typified by an added investment of capital in the building and its productive instruments, beyond which it is not worth while to go at a given time. This is the intensive margin. The extensive margin of land use is not to be found in retail trading or in any other use of city land. There is usually no no-rent land within the limits of the ordinary city.

What occurs is that a business, as it expands, pushes constantly out from the focus of the most desirable land, taking in locations successively less desirable till it reaches those parcels of land which are no more useful for this business than for other uses. Beyond this the business cannot at any given time go, because there now appear other uses for the land which will yield a higher rent, this rent being determined by the next income to be derived from some other use of the land and resting finally upon the prices obtainable for the services of the land in this other—more productive—use. There is equality of return between the intensive margins on all the different parcels of land devoted, at any given time, to the particular business, and as conditions change this equality tends to maintain itself.

The value of land. It is the specific income from land—the economic rent—which, together with the interest rate, determines land values. At an earlier point in our study,[1] we considered the discount process as it applies to the valuation of durable capital instruments. It was there observed that, whereas the demand schedules for capital goods are always derived by discounting expected future income, cost of production operating on the supply side is also a factor in the value of such categories of these goods as are reproducible. In the case of capital instruments which cannot be reproduced, it is the discounting of future income which alone operates on both demand and supply to determine value.

With trifling exceptions of insufficient moment to affect the general rule, land is to be classified as a nonreproducible capital good, and therefore its value varies directly with economic rent and inversely with the interest rate. This is not to say that the sale price of any piece of land is determined by the present rent capitalized at the present rate of interest. Considerations as to probable future changes in rent and the rate of interest always enter into calculations of the present value of future income but alterations

[1] Cf. Chapter XIII.

of land values resulting from these considerations are all in harmony with the basic principle, which may be stated as follows: *The value of land (including all natural agents) is derived, through demand and supply, from buyers' and sellers' estimates of the discounted value of expected economic rent.*

Relation of rent to prices. There is a good deal of popular confusion regarding the causal relation existing between the rent of land and the sale price of the product of the land, since it is commonly believed that high rents cause high prices. But this popular notion confuses cause and effect; it is the price of the product which affects the economic rent of land, not the reverse.

In the case of rural lands, the true relationship between rent and the price of the product becomes quite clear when we review our previous analysis. The price-determining costs of production in agriculture are the marginal costs. But on the margins, as shown in our preceding discussion, cost is equated with price, allowing no producer's surplus or economic rent from the sale of these marginal units of supply. In other words, economic rent is not included among the price-determining marginal costs. Surplus receipts are obtained from the sale of the units of supply other than marginal, because the existing market price makes production profitable beyond the point of diminishing returns. If rents are high at any given time, it is because prices are high.

As to urban lands, the fallacious belief that high rents cause high prices may be traced to the misconception of the nature of the product of these lands. The product of land devoted to retail trading is obviously not the articles offered for sale in the shops. It is an intangible service, the service rendered by the merchant to the social group by promoting the ease and convenience of supplying human wants. If we wish to speak of the price of this service, it should be clear that we refer to the reward of the merchant, not the sale price of his stock in trade.

Any attributes of a site which increase the merchant's gain on a given expenditure of capital and labor in the retail trade beyond the amount received by the users of marginal land will bestow a rent on the site. In some cases a given site will increase the merchant's profit for no other reason than that it places him in a position to charge high prices. Here there is a relation between rent and the selling price of goods, but the causal connection runs from price to rent, not from rent to price. The immediate cause of the

high rent, as always, is the large return of the entrepreneur, but the high return itself is caused in this case by high prices.

But the selling price of his goods is only one of the factors affecting the entrepreneur's net income; another factor of equal importance is his cost of operation. If a merchant has low operating cost because of the character of the site upon which his shop stands, it is the land which really produces his extra profit, and the owner of the land will absorb the excess in the form of a high rent. Among mercantile enterprises, large volume of turnover in comparison with necessary overhead expense is a primary cause of relatively low cost of operation. Sites on crowded thoroughfares and at the transfer points of traction lines permit a volume of business so much in excess of that done by shops in less effective locations that the merchants who possess these desirable sites can gain large returns, at prices no higher than those charged by their competitors. In such cases high rents will be associated with low selling prices of goods, though it is still correct to say that the high rent is conjoined with a high reward for the merchant's service—*i.e.*, with a high income or gain. These illustrations have been drawn from retail trading, but the student will have no difficulty in applying the principle embodied to lands in other uses.

Quasi rent. It will be recalled that the principle of diminishing returns, as we have stated it, is applicable under appropriate conditions to the exploitation of any factor of production. The condition essential to diminishing returns is that one of the factors be held constant while those which coöperate with it are increased. A labor group, if absolutely fixed in number and indispensable to productive enterprise, will derive a surplus income determined by forces analogous to those discussed in the preceding pages. Similarly the owner of capital instruments, if their supply cannot be increased and they are in great demand for productive purposes, can exact a payment of this kind for their use.

As the variable factors increase in amount, the output of the combination, while increasing absolutely, will diminish as measured in proportion to the increasing factors. As we shall see in subsequent chapters, this principle is of great importance in determining the other functional types of income—interest, wages, and profits—as well as rent.

Ordinarily the competitive forces of the market fix the rates of return for labor and capital in a quite different way, because

ordinarily and in the long run the supplies of these factors can be, and are, increased in amount. It should be kept in mind, however, that when a segment of the labor or capital supply is placed in favoring circumstances—limited in amount and combined with increasing supplies of the other factors—its income will diverge from the returns normally allocated to that factor through the market and become a *quasi rent.*

Economic rent distinguished from other income. The condition of fixed supply is much more important in determining the income of land than that of the other factors. There is no productive factor, other than land, which cannot be increased in amount over long-run periods, either in its entirety or in its special forms. Hence the owners of these other factors of production can enjoy surplus incomes from their use only in exceptional circumstances and for short-run periods. But the income from land is normally a surplus income—producer's surplus.

Let us illustrate this distinction by comparing the income from land with the income from capital in a selected situation. Suppose that an office building is erected at a cost of $100,000 upon a certain city lot. A year's operation shows the following results: total income, $15,000; operating costs, $5,000; interest on capital (at 5 per cent), $5,000; net income, $5,000. This last item is true producers' surplus: it is, in fact, economic rent, being due solely to location on this particular piece of land. So long as conditions generally remain unchanged, it will persist.

Suppose now that a shrewd entrepreneur builds, at a cost of $100,000, a ship to carry passengers from the city to a near-by shore resort. The results of his first year's operations are exactly the same as those of the office building; *i.e.,* total income, $15,000; operating costs, $5,000; interest on capital, $5,000; net income, $5,000. Here also is a surplus income, the reward of superior business foresight. But it does not rest on the firm foundation of economic rent. Very soon other entrepreneurs, seeing the extra profits offered, will build other ships. Competition will reduce the rates charged passengers; the gross income per ship will decline in consequence; and finally there will accrue to the owner only the $10,000 necessary to cover operating costs and interest on his capital investment. Thus does competition tend to make short-lived the surplus gain arising on occasion from shrewd applications of capital.

"Unearned increment." Economic rent as a private income may seem hard to justify on grounds consonant with the common man's impulsive sense of justice and fair play. Economic rent is not a payment for any service rendered by the recipient, for it exists only when the recipient's income is in excess of adequate remuneration for all his labor as a worker, all his service as a saver and investor of capital, all his utility to society as an entrepreneur. A service, it is true, is rendered in exchange for economic rent, but it is rendered by nature, not by the rent receiver. In other words, society, in permitting the existence of economic rent as a private income, is making payment for a service which could be had for nothing.

Because rent is of this character, it has often been described—and condemned—as an unearned increment. As the words imply, this name refers particularly to the tendency of rent to increase as population grows. But there is no reason in principle why distinctions should be drawn between different cases of economic rent. Whether rising, falling, or remaining constant, all true economic rent is, in the common meaning of the term, "unearned" by the landowner; it is an addition or increment to his income in excess of adequate payment for every productive service he performs.

It is of course true that the present recipient of economic rent may have invested his savings in the purchase of the land and is therefore receiving only a justifiable return on his investment. For example, assume that someone pays $40,000 for a site which returns an economic rent of $2,000. Clearly there will be no unearned increment in the income of this individual—at least until the economic rent increases—for the yearly rent represents only the normal return on his financial outlay. But the former owner is still receiving $2,000 yearly from the investment of his receipts from the sale of the land; and if we agree that part or all of this income was unearned increment before the sale, it is still unearned so far as he is concerned.

Rebelling against the apparent injustice of the economic rent income, people with a bent toward social reform have frequently proposed that this income be taken out of private hands. These programs of reform have taken many different forms, from outright nationalization of land, as an extreme example, to milder proposals to appropriate a part or all of the economic rent for social purposes.

The method suggested for such appropriation of rent is usually taxation; and this proposal, too, has sometimes taken the extreme form—as in the case of the single-tax movement—of demanding that the state tax all landowners the full amount of the economic rent. The condemnation of rent as a phase of their general condemnation of all inequalities in the distribution of income is an element in the reform programs of the socialists, whose proposals we shall consider in a later chapter.

EXERCISES

1. The figures below show the results of successive added expenditures on different grades of land devoted to the production of wheat:

GRADE A LAND		GRADE B LAND		GRADE C LAND	
Added expenditure	Added product	Added expenditure	Added product	Added expenditure	Added product
$100	200 *bu.*	$100	100 *bu.*	$100	80 *bu.*
100	150	100	90	100	70
100	125	100	80	100	60
100	100	100	70	100	50
100	80	100	60		
100	60	100	50		

 (a) Convert the figures for diminishing physical product with additional outlays on different grades of land into figures expressing rising marginal cost.
 (b) Plot the curves of marginal cost for the production of wheat on the three grades of land (on one chart).
 (c) How far would cultivation be carried on each grade of land, if the market price of wheat were $1.00 per bushel? If the market price were $1.25 per bushel?
 (d) Locate the intensive and extensive margins of cultivation when the price of wheat is $1.00 per bushel; when the price is $1.25.
2. A piece of land occupied by a warehouse is leased for $12,000 per year. Taxes, insurance, maintenance, and depreciation consume $5,000 per year. The building cost $100,000. The rate of interest is 5 per cent. Explain the two elements of which the owner's net income is composed, and determine the amount of each.
3. With the construction of the Erie Canal, the price of wheat in New York City declined. How did this affect the intensive and extensive margins of cultivation in New York and New England and the rent of land devoted to wheat production in this area?
4. Explain how each of the following factors may influence the use of urban sites:
 (a) Convergence of subway lines
 (b) Proximity to schools

(c) Easy access to railroad terminals
(d) Freedom from noise
(e) Proximity to tidewater
(f) Migration of population to suburban areas

5. (a) How would the movement of the New York Stock Exchange to New Jersey affect rents in the New York financial district?
(b) How would the replacement of an elevated street car line by an underground transit line affect rents and the use of urban sites along the route?
(c) How would the development of the St. Lawrence Waterway affect urban rents in Boston? In Chicago? In Baltimore?

CHAPTER XXI

Interest

Introduction. The role played by capital in the productive system was discussed in an earlier chapter,[1] and the process by which capital is accumulated by society was made clear. For our present purpose the essential feature of this process is the time element which is inevitably involved in the provision of capital. In the simplest case, that of a capital instrument constructed by its user, this time element consists of the interval of waiting which must intervene between the expenditure of effort in making the capital instrument and the reward for this effort in terms of income. This time lag may be of either long or short duration, but it is always present. And to the extent that waiting for income is irksome to human beings, this factor will operate to restrain the supply of capital instruments and to prevent their marginal productivity from falling.

In modern society the provision of capital instruments is a more complex process than in the case of the man who takes time out to construct a simple tool for his own use. But behind all these complexities, stands the inescapable fact that society cannot be supplied with capital unless human beings are willing to wait for deferred returns. Laborers cannot be set to work constructing durable goods—vehicles, buildings, ships, railways—unless their needs for consumable goods are met while they are so engaged. To meet these needs some group must provide out of its own income the necessary purchasing power.

This group takes the first step toward capital formation by setting aside portions of its money income. Ordinarily these savings are deposited in the banks and transferred by them to borrowers. On occasion, however, the saver of income may transfer his funds directly to entrepreneurs by buying issues of stocks and bonds in the investment market.

However it is done, the essential feature of the transaction is the

[1] See Chapter II.

transfer of present income by one person or group of persons to another, in anticipation of a reverse transfer at a later date.

As long as waiting is irksome and capital accumulation is done voluntarily, society cannot obtain an adequate supply of capital instruments unless the suppliers are allowed a share of the social income apart from the ordinary cost of producing capital instruments. When we analyze the ordinary loan contract, we discover that it represents a transaction in which present and future incomes are exchanged against each other. The interest element in the loan contract is a premium paid to compensate the lender for his waiting—paid by the borrower who willingly sacrifices a larger amount of future income to acquire the more highly prized present income.

The loan market and the rate of interest. Like any other price, the rate of interest is determined in the market by the forces of demand and supply. The modern world has evolved an organized market in which exchanges of present goods for future goods are constantly taking place; viz., the loan or investment market. Because of the vast number of transactions which take place in the loan market and the innumerable ramifications of this market through the business world, the rate of interest there determined becomes the governor of the terms of all types of transactions involving an exchange of present for future income.

The supply of loan funds: Savings of individuals. The analysis in this chapter will be directed to the supply side of the loan market. Much of the saving by individuals results from a comparison of present sacrifice with future reward, as was implied in previous statements in this chapter. The extent of sacrifice is determined by the individual's rate of time preference; the reward accruing in the future is determined by the interest rate. When the rate of interest is higher than an individual's rate of time preference, he will be induced to save. Saving, however, raises the rate of time preference by impoverishing the present and enriching the future; hence the inducement to save at a given rate of interest will diminish as larger portions of present income are saved.

The amount of loan funds arising from this source tends, therefore, to vary directly with the interest rate. Supply schedules of such funds, beginning with small quantities when the interest rate is zero, will show successively larger quantities offered for loan at rising rates of interest.

There is reason to believe, however, that some saving is done by individuals without much regard to the interest rate. Many people make provision for old age, for the care of dependents, for unpredictable misfortunes, not as a result of the lure of interest but as matter of habit or custom. Such savers have the opportunity to obtain interest on their savings, but it is unlikely that they would abandon habits of thrift and forethought even though no interest were paid. There are in fact multitudes of people who set aside portions of their income under arrangements which preclude the receipt of interest; an example is the practice of the well-to-do of carrying checking accounts at the banks in excess of their actual needs. It is even probable that some people could be made to pay for the privilege of saving under assurances that their savings would not be lost: indeed the banks have at times been able to charge depositors for the safekeeping of their funds.

Corporate saving. It is important to note also that under the operation of our economic system, large sums are continually being saved without the conscious decision, frequently without even the knowledge, of the individuals who might otherwise have the money to spend. One tendency of this order is an out-growth of the corporate form of organization, especially in the large-scale enterprises so common in the modern business world.

Immediate control of the policy of a corporation lies in the hands of the directors, who are in turn responsible for their actions to the stockholders who elect them. One of the problems for which the directors are responsible relates to the disposition of the corporation's net income.

It is never the practice of prudent directors to distribute the entire net income in the form of dividends to the stockholders as rapidly as it accrues. A part is usually set aside as a reserve against future contingencies which may imperil the corporation's ability to pay its debts; another part may be devoted to improving and expanding the material equipment of the business; still another, to experimentation with new processes and selling methods. These decisions to reinvest the corporate income are the result of reason and calculation on the part of the directors, and it is of course the expectation of future returns from the investment that plays the major part in determining the amount saved.

In the case of the large corporation, whose stock is widely scattered among many investors, large and small, it is of course quite

out of the question for the individual to make these decisions. Authority over these matters must of necessity be delegated to the directors, who may thus, in the exercise of their judgment, build up the corporate capital out of current earnings which would otherwise go to the stockholders in dividends.

The activities of the insurance companies afford similar examples of corporate saving. The "level premium policy," the customary form in which life insurance is sold, results in the accumulation of immense funds in the hands of the insurance companies as reserves against their liabilities to the policyholders. There is little reason to believe that the average policyholder adjusts the amount of his contribution to the reserves of insurance companies by scrutinizing the rate of interest.

In view of facts such as these there are some who go so far as to say that there is no correspondence at all between the interest rate and the supply of loan funds. This is clearly an exaggeration. We know that many individuals and corporate bodies do adjust the amount of their savings to the rate of return on loans and investments.

Many people are in a position to be either borrowers or lenders, the choice depending upon the relation of their own rates of time preference to the market rate of interest. In particular the business man may either borrow funds for use in his business or lend capital to others.

Supply of loan funds and the rate of interest. Since so large a part of the total supply of loan funds is provided by savings which are automatic or habitual or involuntary, it is impossible to depict the supply side of the loan market as a series of quantities varying directly and smoothly with variations in the interest rate. Moreover we do not know how large an amount of savings would be forthcoming if there were no interest payment. Society has never confronted that situation and we have no experience upon which to base a decision.

Savers of all sorts, no matter what their motivation, have always been able to obtain interest. It may be pertinent to the question here at issue to point out that the volume of savings has never been large enough to satisfy all the demands for loan funds without requiring the inducement of an interest rate, and that this fact argues the existence of some deterrent to saving which must be overcome by the offer of reward.

We conclude that the total supply of loan funds tends to vary directly, but only roughly and erratically, with the interest rate. Under the conditions prevalent at a given time, it is probable that there is a hypothetical rate of interest which would bring the forces of demand and supply into stable equilibrium; that is, a rate which would evoke only that quantity of loan funds which the borrowers would be willing to take off the market. But since the supply of funds responds sluggishly to changes in the interest rate, there is no assurance that this equilibrium rate will actually result from the interplay of demand and supply.

Over short-run periods, demand exerts the predominant influence on the market, setting the interest rate in accordance with marginal productivity of capital as reflected in the expectations of the borrowers. This rate may or may not clear the market. Over long-run periods, the rate of interest tends, through its delayed influence on the supplies of loan funds, to be so adjusted as to move toward the point which would bring about equilibrium.

The rate of interest and the quantity of money. Each generation brings to the front a group of reformers who, either themselves debtors or impelled by a compassion for the debtor class, propose to abolish interest or to lower the rate of interest by increasing the quantity of currency. Their program is based on the belief that the rate of interest varies inversely with the quantity of money, this belief resting in its turn on the assumption that borrowing constitutes a demand for money and that therefore the price which the borrowers pay (*i.e.*, the rate of interest) must be controlled by the relation of this demand to the supply of money.

This assumption would not be far from the truth if the term "supply" were used in a sense coördinate with the term "borrower's demand"; if, that is to say, it referred to the supply of money available for loans. But we have seen that the borrower does not want money itself, but money's worth. Since the effect of an increase in the quantity of money is to reduce its purchasing power, a larger amount will be required to accomplish the purpose of the borrower. It follows that, as the price level rises in consequence of an increasing supply of money, the demand for loans measured in dollars and cents also rises, thus tending to offset the effect of the increased supply.

Experience may seem to bear out the opinion that the short-term interest rate fluctuates with the amount of money in the country.

The transactions of borrowers are carried on to a large extent with the banks, and their knowledge of the loan market is formed out of their experience with these lending institutions. From this experience they learn that, when the quantity of money in the bank's vaults rises or falls in proportion to loans carried by the bank, the interest rate on loans tends to move in the opposite direction. The banker will explain the low rate of interest which accompanies a relatively large cash reserve by the statement: "Money is plentiful," and the high rate which accompanies the relatively small cash reserve by the statement: "Money is scarce."

But here we have to do, not with the amount of money in the country, but with the amount in the banks and with the motives which induce people to deposit their money as an alternative to spending it. If the bank reserves increase only because of an increase in the total stock of money—that is, if the proportion between the amount of money in the banks and that in circulation remains the same—the greater lending power of the banks will be offset by the increased amount of money required by the borrowers to conduct their transactions on the higher price level. If however the bank reserves increase faster than the amount of money in the country, this is an indication that the people, for one reason or another, are increasing the proportion of money available for loans. In this case there will be a tendency for the interest rate to fall, but obviously the cause of the fall is to be sought in the motives which have induced the people to save more and spend less, not in the amount of money in the country at large. Nor can it be shown that an increase in the total quantity of money will enable the people to save proportionately more than formerly, when account is taken of the rise of the price level which money inflation causes.

The conclusion we have reached has to do with the ultimate effects of changes in the quantity of money. It still remains to inquire how the interest rate is affected by the transition phenomena of the period during which the quantity of money is increasing or decreasing.

We know that when prices are rising under the influence of currency inflation, businessmen strive to take advantage of the favorable market by expanding their enterprises through increased borrowing if, under these conditions, the new stocks of currency are so distributed that the bank reserves temporarily fail to main-

tain their ratio to the amount of money in the country, the need for loans will outstrip the quantity of money available for loans, and the interest rate will rise.

On the other hand, when prices are falling because of a contraction of the currency, money may find its way into the banks in unusual amounts, owing to the tendency of the people to restrict their expenditure during periods of depression. At the same time, the general feeling of business uncertainty will tend to reduce the demand for loans; hence the interest rate may fall when the quantity of currency is decreasing. The effect on the interest rate, in other words, will be the opposite of that expected by those who believe that the rate of interest varies inversely with the quantity of money.

We should note also that the effect of a rising price level is to reduce the real worth to the lender of the principal of loans which he made prior to the rise of prices. A loan of a thousand dollars loses one-half of its purchasing power if the price level doubles before the loan is repaid. Rising prices therefore penalize lenders in the case of all pre-existing loan contracts and benefit the borrowers to an equivalent extent. Conversely, a falling price level impoverishes borrowers and enriches lenders.

It is to be expected that borrowers and lenders will take account of these effects of a fluctuating price level and attempt to guard against them by raising or lowering the interest rate. During a period of rising prices, the interest rate tends to rise as an offset to the depreciation of the principal of the loan; during a period of falling prices, it tends to fall as an offset to the appreciation of the principal. This tendency coöperates with the one just discussed to cause a rising interest rate when the supply of currency is increasing, and vice versa. These, to repeat, are transition phenomena. They disappear when the price level is again stabilized.

Lack of uniformity in interest rates. Thus far we have spoken of the interest rate as if there were a single uniform price applying to all types of transactions in the loan market. But when we turn to the business world for illustration of our theory, we are impressed at once by the absence of any such uniform price. Different rates of interest appear in the market at the same time, each rate applying to a particular type of loan transaction. For example the following schedule of rates was published on three different dates for transactions in the New York money market:

Type of Loan	PER CENT		
	July 26, 1929	Feb. 16, 1935	Jan. 30, 1948
Call loans	10	1	1½
60–90 day loans	8	¾ and 1	1¼–1½
Commercial paper			
Best names	6	¾	1¼–1⅜
Other names	6	1½	1½

Some of the variations appearing in the first item of this table are due to conditions peculiar to the banking business. A large part of the banker's liability consists in deposits payable on demand, against which he must maintain assets in liquid form. Bankers in a money center like New York receive from banks in inland cities large deposits of cash subject to continuous withdrawal and replenishment. For obvious reasons it is not profitable for these bankers to hold such deposits in the form of idle cash. The call loan is the best form of investment for funds of this kind; it is resolvable into cash on demand and secured against loss by a deposit of collateral. For this reason the call loan in a large money center commands a rate of interest which fluctuates sharply in response to variations in the amount of money seeking this particular form of investment.

The student will realize that the interest rates listed in the second and third columns of the table indicate an abnormal condition of the money market. Prolonged business depression had caused the banks to carry an unusually large portion of their assets in cash. This abnormal supply of short-term loan funds, coupled with the decline of demand in the business world and the government "cheap money" policy, resulted, on the dates given, in extremely low rates of interest in the New York money market.

True and nominal interest rates. On loan transactions not subject to these exigencies of the banking business, variations in the interest rate are often more apparent than real. We must view interest as a payment for a single kind of service; *i.e.*, the service of postponing income by a voluntary exchange of present goods for future. If the man who renders this service performs at the same time a number of other functions and receives a single payment as a recompense for all of them, only a part of the payment, strictly speaking, can be called interest; the remaining part is composed of as many different elements as there are different functions to be rewarded by the payment. Now it happens that the term "interest" is commonly used in the business world to define just

such a composite payment, of which only one element is true interest.

The easiest of these elements to recognize is the insurance premium, which generally accompanies every interest charge. When dealing with true interest we must assume that the lender is assured of his future return; but as a matter of fact no loan is absolutely secure. When repayment of a loan is uncertain to a marked degree, the fact evidences itself in an apparently higher interest rate. This interest rate consists of at least two elements; *i.e.,* interest itself and a premium for risk and uncertainty.

There are also other elements commonly included in commercial interest rates which do not correspond to the true interest concept. The rate which a bank charges its customers, for example, is not merely a charge for waiting; it is also a charge for consultation and other forms of service. Interest paid a stock or investment broker also carries a service charge of similar nature.

The most common occasion for confusion of interest with other forms of remuneration is the use of an entrepreneur's own funds in his own undertaking. In any such case it is exceedingly difficult to segregate true interest from wages which the entrepreneur has earned by his own personal efforts; *i.e.,* the wages of management.

Moreover what is in reality merely an "eating up of principal" is sometimes falsely regarded as interest. No true interest ever accrues from the use of a capital instrument unless provision has been made to replace it as it gradually wears out, and yet interest is often confused with depreciation.

In summary of these points we may say that nominal and true rates of interest never coincide unless the investment is one in which the investor bears no risk, exercises no managerial powers, and encounters no depreciation in connection with his income. Very probably no investment can be found which measures up exactly to this ideal, but it is approximated in the long-term bonds of a stable government, returns on which may be accepted as a fair exemplar of true interest.

Changes in the rate of interest. Over extended long-run periods, a remarkable stability in the rate of interest is to be observed. When in the seventeenth century the interest rate in Holland fell to 3 per cent, it stood at a point only slightly lower than during the flourishing days of the Roman Empire. In the following century

the rate rose slightly, only to fall during the nineteenth century to the same low point. Just prior to the First World War the rate for first-grade securities stood again at about 3 per cent.

But this stability over long ranges of time contrasts with fairly sharp fluctuations in short-run periods. The rate for current loans is a market price which tends to respond to momentary changes in the factors of demand and supply. There is also a tendency for the rate to rise in the wake of rising prices during the "prosperity" phase of the business cycle and to fall to a low point during the subsequent reaction.

But these responses of the interest rate are at best sluggish and retarded in action, a fact which has unfortunate economic consequences. The rates of pre-existing loans, being determined by legal contract, do not change at all in response to fluctuations of the business cycle or change in a disjointed manner at the end of the respective loan contracts. Moreover the rates on current loans are so influenced by customary business practice and other forces interfering with perfect competition in the money market that they too fail to move promptly and smoothly in harmony with changing business conditions. If the price of loans were to rise promptly during the early stages of rising prices, it would act as a brake on unhealthy credit expansion and thus tend to prevent a subsequent crisis, but because of these market frictions it fails to keep pace with the situation.

Similar lagging of the interest rate may be observed during the early stages of the business depression which follows a crisis. The general collapse of prices at such a time reduces the value of business assets and current business income, while the sluggish behavior of the interest rate prevents a corresponding reduction of business indebtedness and business costs. This tends to undermine the solvency of enterprises which are in debt as well as to make it more difficult to place current business transactions on a profitable basis.

If the depression is prolonged, the rate may fall to abnormally low levels, as in 1940, because of the continued lack of demand from the business world. This lack of substantial return from loan funds exerts a disturbing influence on the economic system through its effects on the banking structure and on all enterprises and individuals dependent on income from invested funds.

Interest as a private income. That owners of capital possess a right to share in the distribution of the product of industry is due to certain essential characteristics of our type of economic order. Our civilization is founded on the use of great masses of capital; progress in human well-being requires that we set aside each year an enormous amount of present income for the repair and replacement of existing capital instruments and the addition of new ones.

The industrial mode of production calls for two distinct kinds of service, working and waiting. Neither of these services by itself can maintain the flow of consumable income at a volume sufficient to satisfy the wants of men on the present plane of living. Man's labor of hand or brain, unless reinforced by his willingness to wait and to endure the costs of waiting, must be applied to direct production without the help of those product-multiplying aids which we call capital instruments.

Although the service of waiting is an indispensable productive function, this service is not rendered at the time the interest income is received, but was performed in the past when the saving was done. Indeed in many cases the service of waiting is not even performed by the same man who receives the interest, but by some predecessor from whom he has inherited his wealth. One consequence, therefore, is the existence of certain people who appear to enjoy the right of sharing in the income fund of society while making no concurrent contribution to that fund. This spectacle affronts the sense of justice of many a man who does not reason clearly with regard to the capitalist's function in society.

Other grievances associated with the existence of interest as a private income give impulse to this attack upon the institution. Although interest income is, as a matter of fact, widely dispensed among millions of men, women, and children of modest wealth, it occasionally occurs that, through inheritance, large fortunes based on interest have accumulated, permitting certain persons to live in idleness in the midst of a busy world. Even though their fortunes may be only the reward for past saving, either by themselves or by their predecessors, their present situation instills in some people a sense of jealousy or grievance. Reaction against the institution of interest as a private income leads to organized movements which propose a thoroughgoing change in the present economic structure of society.

Regulation of interest. In part, the modern attack upon interest is an inheritance from an earlier day and a simpler stage of societal development. Throughout recorded history there have been individuals and communities which have condemned the receipt of interest by lenders as an immoral and indefensible practice and have taken steps to abolish or strictly regulate it. Perhaps the best known example of this attitude is supplied by the Church during the Middle Ages. For over a thousand years, indeed down to the thirteenth century, the Church Fathers carried on an unceasing attack upon interest. By the canon law its receipt was strictly forbidden to Christians on pain of spiritual punishment in this world and the next.

These laws became obsolete and gradually disappeared in all modern nations. But vestiges of them remain in the form of "usury" laws, and the point of view which they represent comes to the front in modern times, not only as a phase of the program of the socialists, but also in the form of sporadic political movements among more conservative men.

There was some ground for moral condemnation of the interest income in early stages of social development. Almost all loans in that day were loans for consumption. The spendthrift borrowed to enlarge his supply of consumable goods; the necessitous man, the widow, the victim of misfortune were compelled to borrow to eke out a day-to-day existence. The king or the lord of the manor borrowed to cover the expenses of a war or crusade, or to heap favors upon the court favorite of the day.

Such loans were indeed "barren," to adopt the expression of the moralists of the time. Certainly such a borrower was not enabled to bear the charge which interest laid against his future by creating income through his use of the borrowed funds, as does the entrepreneur-borrower today. And in the common run of cases, the position of the creditor gave him the appearance of exploiting the necessities of his brother man. The usury laws of the time were passed in defense of the weaker man, the debtor, to protect him from being surrendered by misfortune into the hand of the voracious money-lender.

Where this type of loan has carried over into modern society, we continue, with general approval, to regulate the interest rate out of concern for the borrower. Though interest is no longer prohibited

on loans for consumption, all modern countries make some provision for the protection of the most necessitous borrowers from the worst consequences of their thriftlessness. In America the states regulate the pawnshops by laws fixing a maximum interest rate and requiring accessible records of all transactions.

But the typical debtor of today is not the timid victim of misfortune, but the most able, most self-reliant of us all, the business entrepreneur. Borrowing for production in the normal course of events places the borrower in position to meet the interest charge without trenching on the income which otherwise would accrue to him. However, despite the fact that the day has passed when spendthrift loans comprised the major part of the business of the loan market, the antipathy toward interest, especially a high rate of interest, remains and finds expression in laws intended to restrict the rise of the rate in the financial world.

Recognizing that exceptional cases arise which require exceptional treatment, it should be evident that such laws are in general both futile and injurious. They are futile because, when conditions arise which enable the entrepreneur-borrower to pay high rates of interest and therefore create in him a disposition to do so if necessary, the purpose of usury laws will be defeated by connivance of both borrower and lender.

They are injurious, both from the standpoint of the individual borrower and from that of society as a whole, for if any branch of business or any geographical locality is handicapped by lack of capital (and this is invariably the case when complaints are heard about the high rate of interest), it is obviously to the advantage of all entrepreneurs to facilitate an increase in the supply of loan funds. The basic cause of the condition cannot be removed by arbitrarily relieving present debtors of a part of their interest costs.

What is needed is renewed inducement to saving and investment. Usury laws, by depriving savers of a part of their inducement to save and by injecting into the loan market the incalculable element of political interference, tend to perpetuate the evil which lies at the root of the protest; *i.e.*, a paucity of loan funds. Not only the individual who is attempting to carry on productive enterprise under the handicap of small capital supply, but the entire community, which shares in the rewards of the capitalistic process, is benefited by allowing the interest rate to fluctuate freely in accordance with the forces of demand and supply.

"Over-saving" or "under-investment." It was made clear in a previous chapter that money incomes of all sorts arise from the process of producing goods for sale on the market. The recipient of an income—be it wages, interest, or rent—is receiving payment for a contribution which he, or productive instruments owned by him, has made to the value of a product. These payments are made by business managers in expectation that the product will command a market price sufficient to reimburse them. For the productive system as a whole to function smoothly it is necessary that all these money incomes be spent in the purchase of goods and services; in other words, that they should be used *in toto* to take off the market the goods whose expected sale gave rise to them. If the sum total of money incomes is so utilized, the market will be cleared, and the economic system can continue to operate at the existing volume of production.

This statement does not imply that it will be injurious for individuals to save a portion of their incomes. When the economic system is functioning aright, money saved by one individual is spent by someone else; that is, by an entrepreneur to whom savings are transferred through the investment mechanism of the country. In this situation savings are also spent, and the net effect is merely to create a demand for different kinds of goods than would have been bought had all individuals spent their entire money incomes and saved nothing.

The danger in the situation is that savings may not be transformed into investments and hence may not appear as purchasing power in the market. If this should happen, the savings of individuals would become hoards of unutilized purchasing power; and the money income of the nation's citizens would not be applied *in toto* to clear the markets of goods and services and thus keep the productive system in full operation.

Such a situation is better described as "underinvestment" than oversaving. What is really happening is that new investment outlets are failing to keep pace with the volume of savings; for some reason or combination of reasons, the investment market is failing to absorb the unspent incomes of individuals. If this maladjustment reaches serious proportions, a chain of events may follow: retrenchment of productive activity; decline of demand for labor, land, and capital; collapse of the rates of wages, interest, and rent: in short, business depression, with its concomitant effects of unemployment and declining incomes.

The absence of a close relationship between the supply of loan funds and the rate of interest accentuates the danger of maladjustment of savings to investment. If the volume of savings were closely controlled by the interest rate, there would be no possibility of prolonged "underinvestment" for then an excess supply of loan funds would cause a decline of the interest rate, which in turn would cause the volume of savings to diminish until equilibrium was re-established. But, as we have seen, a considerable part of the total savings is controlled by motives and practices which have little regard to the rate of interest; hence maladjustment between saving and investment is possible.

CHAPTER XXII

Wages

Approach to the problem of wage rates. A study of the forces which determine the rate of wages presents peculiar difficulties. The great mass of working people is divided into a bewildering number of groups and subgroups, each receiving a different rate of income, and within these separate groups the earnings of individuals will be found to vary in such fashion that any attempt to formulate a general law governing the rate of wages appears hopeless at the outset.

In view of these complexities, the first step must be to simplify the problem. This can be done first by limiting attention at the start to some group within the great class of wage earners whose members are fairly homogeneous in quality; that is to say, a group so alike in function and capabilities that one man can be substituted for another without material loss to the industry. In the second place, we shall assume a condition of free competition and a perfect labor market and develop the theoretical principles which would operate under such confessedly artificial conditions.

Under these assumptions, there will tend to be a uniform wage rate for the group, determined by demand and supply. The qualifications required to adjust this simplified picture to the complex facts of the industrial world can be made after we have analyzed the forces which tend toward a uniform wage rate within this section of the labor market.

The demand for labor. It is necessary here only to recapitulate statements made in a previous chapter. We have seen that the determining influence on the demand side of the labor market is exerted by profit-seeking employers who voluntarily undertake to give the workers a money income in exchange for their services. Each employer expects the workers to create the income out of which their wages are paid, by increasing the revenues of the enterprise by an amount at least equal to the wage charge. In other words, the employer's demand for labor arises out of the expected monetary productivity of labor.

In any homogeneous labor group, it is not the total, but the marginal, product which governs the employers' demand for labor. This is because employers are not confronted with the choice of hiring the entire labor force as a single unit, or hiring none at all. They employ workers as individuals and are concerned with the specific contribution of the individual to the income of the enterprise. This marginal productivity of the labor group is defined as *the amount per laborer by which a small change in the number of laborers changes the total product.*

The productivity of labor conforms to the law of diminishing marginal productivity. No employer can go on expanding his labor force, other factors of production being constant, without noting a decline in the amount of product attributable to one unit of labor. Since this is true, an employer, given a rate of wages, will tend to increase his labor force until the marginal product is no greater than the wage rate. Evidently the lower the rate, the greater the number that can be employed before this limit is reached; hence the demand for labor can be expressed in general terms as follows: *The number of laborers demanded by a given employer or by all employers under any given set of circumstances varies inversely with the rate of wages.*

One important technical feature remains to be considered. The workers will want their pay in advance of their employer's sale of the product. Generally at the end of each week the laborers will ask to be paid for the services performed up to that time. But in the majority of cases the products upon which they have been working are not in a completed form and ready for the market at the time the wages are paid. This phase of the matter was touched upon in our study of interest, where the capitalistic process was described as a succession of advances to workers.

But the employer will not thus make advances to laborers unless he is paid for his waiting; that is, unless he receives interest upon the advance payments. The wages which the employer can afford to pay the marginal worker therefore will fall short of the value of his contribution to the product of the business by the amount of the interest charge. Thus the employer is compelled to discount the marginal product of labor in deciding what wage he can offer.

The supply of labor. Other things being equal, the quantity of any good which will be offered for sale varies directly with the price. In general this is true of the supply of labor within any

single occupational group and within the wage class as a whole; but it should be noted that the circumstances back of the supply of labor are in many respects peculiar.

The amount of labor available is a comparatively fixed quantity, not susceptible to much variation in response to changes in the rate of wages. Changes in the total number of wage earners take place fairly slowly, and workers are not easily drawn from one functional group into another. The effect of these impediments is to make the supply of labor comparatively inelastic.

The general law of wages. The forces of demand and supply would, if unimpeded, determine a market rate of wages in a manner similar to that by which other prices are determined in a competitive market. The demand for each type of labor is a schedule of rates representing the discounted marginal utilities of labor groups of differing sizes. Because of its inelasticity, the supply of labor is, at any given time, virtually a fixed number of workers in each homogeneous labor group. The conditions of demand will fix the highest rate at which this number of workers can be absorbed in employment under the free play of economic forces. The general law of money wages may be stated as follows: *The normal rate of money wages for any uniform group of workers is equal to the discounted marginal product of that group.*

Although the individual employer assumes that the wage rate is fixed by market forces extraneous to himself, he plays a part in causing this rate to be what it is. Were he to discontinue his enterprise and withdraw from the market, the workers whom he released would have to seek employment elsewhere, thereby (in theory) lowering the marginal productivity in other enterprises and reducing the wage rate. On the other hand, if this employer is enabled by some development in his business to increase his labor force without reducing its marginal worth, his increased demand for labor can be satisfied only by diverting wage earners from other industries, thus raising marginal productivity in general and increasing the wage rate.

Limitations: Measurement of marginal productivity. The general law of wages is obviously an abstract statement of a tendency, rather than an explanation of all the diversities which may be observed in the business world. We must now add qualifications in order to take account of some of the more important practical considerations.

It is obvious, in the first place, that the appearance of precision given in our illustration to the employers' calculation of the worker's marginal productivity is artificial. In the complexities of modern industry with its elaborate division of labor, no one can say exactly what one worker contributes to the corporate income. At best, these calculations are intelligent guesses based on experiment and experience.

Interdependence of different types of labor. In the second place, it never works out that the different entrepreneurs enlarge their several labor forces to the point where the marginal productivities of similar grades of labor are precisely equal. The number of workers of a certain type which can be employed profitably in a given industry is affected, not only by the marginal productivity of these workers, but also by the necessity of maintaining an effective proportion between the different types of labor in the enterprise.

Complex division of labor creates a situation in which no single type of labor can be employed without necessitating the employment of other types. For example, an industry employing one bookkeeper might not have need for the services of another at any rate of wages, so long as its present scale of operations remained unchanged. To make it possible for a second bookkeeper to produce anything might require the employment of a dozen machinists, a foreman, a shipping clerk, and two or three other types of labor, none of which could be employed profitably at their prevailing wage rates. In a certain large factory producing tennis shoes, one inspector is employed in conjunction with twenty-five workers in other lines; one cutter works with half a dozen stitchers, etc.

These examples illustrate the fact that the demand of any employer for a given type of labor is limited by his force of coöperating types, which is limited in turn by the wage rates of these other labor groups. He tends to increase the number of workers in each group up to the point where their marginal productivity equates with the wage rate, but he may stop short of this point because of his inability profitably to expand the scale of his industry. Obviously the small business is much more affected by this consideration than is the large one, which, with adequate supplies of capital, can take more complete advantage of the gains inherent in complex division of labor. When however attention is fixed on a single occupational group, it is probably true that in no individual

enterprise could there be found that precise equivalence of marginal productivity with the wage rate which our law of wages implies.

The limits of bargaining. As the law of wages has been stated, it would appear that the rate of wages tends to settle automatically and without friction at a point which measures the marginal productivity of the workers. This is because conditions of perfect competition have thus far been assumed. But in practice the rate for any group of wage earners in any given industry is settled by bargain between employer and worker and fluctuates according to the relative bargaining power of these two parties to the wage contract.

The actual rate in any industry, or for any section of the laboring population, may arrive at a lower point than that determined by economic law; on the other hand, it may rise above this point for short-run periods in isolated enterprises or localities. We cannot undertake to follow out in detail these complex variations, but we should ascertain as nearly as can be the limits of this bargaining area within which actual wage rates fluctuate.

Wages above marginal product. Circumstances may arise in which the employer will voluntarily pay some of his workmen more than they are worth, thus leaving himself worse off than if he had refused to hire them at all. Employers are not devoid of feelings of compassion and do at times mix charity with business. Furthermore it may be to the positive advantage of an employer to accept a loss on the wage contract for a short time in order to avoid a greater loss involved in dismissing his employees. During periods of depression, the entrepreneur will look forward to a recovery of the market which will set his industry in full operation again. By keeping an experienced labor force together, he will avoid the considerable loss involved in hiring new workers, breaking them in to the routine of the plant, and training them in his methods of management. It may be worth his while to do this even at the expense of paying his workers for doing nothing. These are but indications of the motives which may impel employers voluntarily to pay wages above the rate set by economic law.

It is also possible for the workers, under certain conditions, to force an employer against his will to pay them more than their marginal worth. By threat of strike, a group of workers which holds a temporary monopoly on the supply of labor may compel the

employer to make a choice between the alternatives of paying the members of this group more than the value of their marginal product, or of suspending operations.

If an employer could always close his plant without loss, it is not probable that he would ever accede to such demands. But this he cannot do. To suspend operations will at least wipe out his own wage income—the reward of his labor of management—along with the possibility of profits. If compelled to do so, he may pay wages above the marginal worth of the workers, surrendering to them a part of his own income up to the point where his share in the revenues of the enterprise has disappeared.

Even here we do not have a final limit to the advance of wages, for it is conceivable that aggressive bargaining by the workers may raise the rate so high that the entrepreneur will not only fail to receive payment for his own labor but will actually operate at a loss. Most employers have fixed costs which will run on unchanged even if they close down their plants. It is the lesser of two evils to run at a current loss, provided the loss is not so large as the total of the current fixed charges.

This condition is of course self-limiting; its ultimate limit will be reached when the employer's equity in the business has been wiped out. But it is quite possible that an employer can be made to bear a loss of this kind, and there are plenty of cases of employers who have been forced into bankruptcy as the result of a series of losses on the wage contract.

We see then that there are three resistance points to a rise of wages. The first and most important is set by marginal productivity. It is rarely "good business" to pay a higher rate than this, and no employer will do so intentionally, unless it be as a temporary expedient in abnormal circumstances. It is also rarely good business for laborers to demand more, since their employment is determined ultimately by profitable enterprise. The second check on rising wages is exerted at the point where the employer's profit and wages of management disappear. This is the maximum limit to the range of bargaining, except in industries materially affected by fixed costs. The highest conceivable rate under any conditions is set at the point which causes a current operating loss equal to the loss involved in closing down the industry.

Wages will rise above the first of these points, that determined by marginal productivity, only when the workers are possessed of some

unusual bargaining advantage. Even then such a rise of wages can appear only in isolated cases and for short-run periods. When his wage cost rises to the second of our resistance points, the employer will soon give up the struggle. No employer will be content to work indefinitely without reward. The duration of a wage rate at a level between the second and third resistance points is automatically self-limiting by inevitable insolvency. We may safely conclude that advances of the wage rate above the point of the discounted marginal product will appear only as exceptional and short-lived occurrences.

Perhaps notice should be given to a type of situation sometimes encountered in the business world which is often cited in refutation of this statement as to the temporary nature of wage rates above the marginal productivity level. Tightly formed and aggressive unions do sometimes raise their wage rates to extravagant heights and hold them there over considerable periods of time. But they do this by monopolizing and restricting the supply of labor. This raises its marginal productivity and drives up the price of the product on the market. The employer who pays these high rates does not necessarily pay more than the discounted value of the marginal product in the monopolistic situation which exists.

Wages below marginal product. For somewhat different reasons, the rate of wages may under certain circumstances fall below the marginal product of the labor group. The belief that employers will be required to pay wages equal to the full economic worth of the hired worker rests upon the assumption that competition works as freely on one side of the labor market as on the other. But this is untrue in actual life. For wage earners obviously are not as a rule in position to take full advantage of competition among employers. To do so, they must be able to play off one employer against his competitors, to move about at will from one locality to another, to discard service in one industry and undertake employment in a different one whenever it is to their immediate economic interest to do so.

All of this requires a degree of mobility in the ranks of labor which the average workman does not possess and which he is prevented from acquiring by (to mention but one factor among many) the mere cost of transportation for himself, his family, and his household effects. It must be remembered that the labor market ordinarily requires the workman to carry his labor to the employer.

Perfect competition in the labor market then implies physical mobility among laborers, to attain which they would have to overcome obstacles of financial expense, ties of home and community, custom, tradition, sentiment.

There seems to be no doubt that the bargaining weakness of certain workers has actually resulted in a wage rate lower than their marginal economic worth. To bring this to pass, the local situation must be one in which the play of ordinary competitive forces is interrupted. This condition of weakened competition may enable employers to obtain labor for less than it is worth in terms of marginal productivity. Such a state of affairs, in particular situations, may be long continued. Conditions of this kind have sometimes been disclosed when wages have been raised by law without causing either unemployment or a higher price for the product.

It would aid us to limit the range through which the wage may fall if we could ascertain with precision lower limits to the sphere of bargaining comparable to the resistance points which check a rise of wages. In a general way two such resistance points can be marked out.

The first of these is determined by the wage earners' traditional standard of living. Among the skilled labor groups especially, self-respecting workers will reject offers of employment at rates which spell a degradation of their standard of life and a loss of social status.

A second and ultimate lower limit is reached when the rate falls to the minimum of subsistence, meaning the income level necessary to support and perpetuate the laboring population. Even if they were content to work for less than subsistence wages, the duration of such a wage rate would be limited by its inability to maintain the supply of labor.

Variations in wage rates: Individual variations. For purposes of exposition, we have thus far assumed that the workers in each specific line of occupation are so homogeneous that one can be substituted for another at will. Under this assumption all members of a given occupational group would receive the same rate of wages. All carpenters would be paid alike, all stenographers, all machinists, all schoolteachers, etc. Everyone knows that this is not strictly true in fact. The wage rate varies for different members within an occupational group and also for different localities.

Of course we are here considering not merely the money payment, but the real wage. It is necessary to be certain that apparent differences in wages are real differences. If there are wide discrepancies in living costs between different localities, variations in money wages may not measure the real incomes of the workers. If employment is erratic, as in seasonal industries, a high daily rate may not mean large annual earnings.

Considering now the lack of uniformity in the real earnings of different members of the same occupational group, it is obvious that this fact may sometimes be due to relative difference in bargaining power such as we have discussed above. Women are habitually paid less for the same service than men. Negro labor in the United States is similarly underpaid in comparison with white labor, and the same is true of recent immigrants in comparison with workers who have acquired American culture and standard of living. These are examples of low wages due to bargaining weakness.

But of chief importance in explaining lack of uniformity in the earnings of different members of the same labor group is the fact that no labor group is quite homogeneous in productivity. If we select any occupation at random, we will observe that all its members do not have the same productive efficiency. Noticeable differences appear in the number of tons of coal credited to the account of different miners at the end of an eight-hour day. The tenders of machines in a cotton mill will differ somewhat in the speed with which they perform their operations and in their ability to sustain a given rate of production through the hours of the working day.

Closer observation will show that these differences in productivity are greater in the case of some labor groups than others. In the crudest forms of labor, the difference in the output of one member of the working force as compared with another is usually slight. So also with the great army of workers in the factories; there is a marked uniformity in productivity among the millions of machine operatives in the great industries. The differences are great among skilled craftsmen, where native ability and training play a relatively large part in the mastering of the trade. And differences are greatest of all in the highest labor groups, the learned professions, and the superior grades of managers and superintendents.

Geographical variations in wage rates. Wage rates differ not only for different members of a given occupational group in the same

place, but also between localities. To some extent, these differences may be attributed to the effect of the bargaining factor which we have discussed. For various reasons, the bargain which fixes the actual wage rate within the limits previously described terminates variously in different localities.

Entrepreneurs are not alike in their efficiency as managers, and consequently different business concerns vary in profitableness. Some units of a given branch of industry are equipped with better machinery or possessed of more favorable markets or otherwise favored as compared with competing units. These differences affect the ability of an enterprise to pay wages, and they also affect the aggressiveness with which the employer bargains for a saving of cost in the wage contract. Employers with the smallest profit and those operating without profit or at a loss are proverbially the most aggressive bargainers.

On the other side of the labor market, differences in the character of the working people, in their standards of life, in the degree to which they are unionized, and in other local conditions materially affect the bargaining strength of different groups of laborers. Such differences in local conditions modify the resistance points in the fluctuation of wages and cause local variations from the norm.

The basic cause of local variations, however, is the fact that the labor market is broken up into small fragments, between which the forces of competition either do not operate at all or operate haltingly and with much friction. While prices of commodities in general in different places cannot differ by more than the cost of transportation between these places, labor is an exception to this rule on account of its extreme immobility. If all the workers in a given occupation could and would move at will over the whole industrial field, local wage rates would tend toward much greater uniformity. But in reality each business center forms a labor market of its own, where the relation of supply to demand may be quite different from that of other centers and where consequently the marginal productivity of the workers will also be different.

Local wage rates tend to vary in rough proportion to the cost of living; they are lower in small towns than in the metropolis, in southern than northern cities, in the Middle West than on the seaboard. This fact lends support to an erroneous theory that the wage rate is determined by living costs. But local differences in the cost of living can affect wages only through their influence

on the distribution of the labor supply and on the bargaining power of the workers.

Self-employed workers. The foregoing law of wages was developed with reference to the hired worker, but its conclusions will serve to explain the money incomes of those who are their own employers.

Here the typical case is complicated by the fact that the independent worker must almost of necessity be an owner of capital and to some extent a manager of enterprise. The physician, the lawyer, the barber—examples could be multiplied indefinitely—possess productive capital of one sort or another and often land as well; their money incomes are a mixture of rent, interest, and wages, no single element in the income being clearly defined in contrast with the rest. It is not impossible, however, to arrive at a rough measurement of that part of their income which is attributable to labor, by deducting allowances for capital and land at the market rates of interest and rent.

In the case of independent workers, no intermediary stands between them and the market. Their money incomes are drawn directly from the buyers of their products. Their incomes rise with a rise in the price of their product, and fall as prices fall, tending to settle at any given time at points which measure the marginal utilities of their services to consumers. The fact that these groups are less homogeneous than are groups of hired workers, that there are, for example, relatively large differences in the productive efficiency of physicians, attorneys, and consulting engineers, tends to bring about marked variations in earnings within these groups.

But apart from these variations in money wages, these classes of the laboring population constitute no real exception to the law of wages. The normal income of the doctor will be high or low in harmony with the relative scarcity or abundance of the services of this profession or, to use the terms of our law of wages, in harmony with the marginal productivity of the group. So it is with other classes of independent workers.

Nonbusiness workers. The more apparent exceptions to the law are exhibited by those groups of workers whose incomes are determined in some material degree by noneconomic considerations. Such are those connected with charitable enterprises, members of religious orders, to a less degree teachers and government officials. There can be no direct relation between the earnings of such workers and the market prices of their products, for the sufficient

reason that the prices are not determined by market considerations. But even here the economic law of wages is by no means without effect.

These occupations must compete for workers with opportunities in the general world of business and professions. The rewards offered cannot be set too far below those afforded by competing occupations to men and women of the same ability, else the former enterprises will be unable to obtain the requisite supplies of labor. The periodic complaint over the dearth of teachers, ministers, soldiers in the standing army, and "good" men in all branches of the civil service shows what happens when the wage policies of these noncommercial enterprises are out of line with the general economic conditions of the times. Making allowance for the nonmaterial rewards of such types of labor as we have been considering, such as prestige, security of tenure, relief from strain of competition, congenial surroundings and associates, and the like, we shall find that wages in these occupations show rough similarities to the ruling rates fixed by economic forces in comparable lines in the business world.

Importance of discounted marginal productivity. Because of the complexity of the labor market, it has been necessary to introduce various qualifications into our law of wages in order to bring it into rough conformity with the facts of industry. We have recognized that the actual wage rate is determined by bargaining and have marked out the limits of the sphere within which the incidence of the bargain will fall. We have taken account of the existing lack of uniformity in the wage rate for members of a given occupational group and have accounted for this by referring to differences in bargaining power and especially to differences in individual productivity. We have also noted local variations in wage rates for a given type of labor and have shown that these result from local differences in relative bargaining power and especially from maladjustment in the labor supply.

But this recognition of complicating factors must not be permitted to obscure the essential truth embodied in the economic law of wages. The wage bargain revolves around the point of discounted marginal productivity as a norm; actual wage rates cannot long persist above this point, and all the developmental tendencies in our society work against their falling far below this point, or remaining for long below it. Homogeneity of labor groups is pro-

moted by various tendencies which we shall consider further on: other forces are broadening the scope of the labor market and promoting a more even adjustment of the local supply of labor to the demand for it.

We may conclude that no alterations in the method and procedure of the labor market can materially affect the fact that the laboring class, in a society founded on private property and individual enterprise, will normally receive wages approximating the value of the specific contribution of this class to society's income. If the laborers are members of groups whose individuals have much the same rate of productivity, the value of each man's contribution will be equated, through the principle of substitution, to that of the marginal man in the group. If they must receive their share of the social income before the sale of their product, their money incomes will be reduced by the discount process through which present and future values are equated. Upon these valid principles rests the economic law of general money wages, the law of the discounted marginal product of labor.

Wages and the contrast of social classes. The greatest differences in wages which we observe in actual life mark the contrast in the economic positions of the several non-competing social classes; *e.g.*, the class of unskilled manual laborers and that of the executives, the managers, the self-employed entrepreneurs. These great inequalities do not measure the irksomeness of the task performed, for often the man who does the most exhausting and unpleasant work is more poorly paid than any other. Nor do they indicate the relative social importance of the type of work as such. It is quite impossible to say whether the managerial function as such is more important than manual labor; both are indispensable.

What the inequalities do measure is the relative marginal productivity of the social groups in question, and this in turn is a function of the relative plenty or scarcity of numbers in the group. The underlying reason for these inequalities is that men cannot transfer at will from the overcrowded to the less crowded social classes; or, to state it otherwise, that the members of these diverse classes do not compete across class lines.

Speaking broadly, there are two contrasting explanations offered for the persistence of social stratification through the ages. There is the belief that conditions of birth and the heritage from his ancestors determine the rank which a man will take among his fel-

low men. Those in the humblest station are predestined at birth to occupy this position in the lowest income group. Those who rise to higher levels do so by virtue of inborn characteristics which can neither be produced intentionally nor balked of their reward when present. With these writers heredity is the controlling factor in the destiny of men; the present stratification of society is a natural order determined by nature's laws.

In contrast with this view stand the environmentalists, who assert that man's social destiny is determined after birth by the forces which play upon his plastic consciousness during his early years. In their opinion the children of the lowest group of wage earners fail to rise above their fathers' station in life because of the effects of the unfavorable environment of their childhood. Able men, so this view goes, are born in about equal proportions in all ranks of society, but those in the lowest ranks are denied opportunities for self-development, opportunities which are enjoyed by the children of the rich, regardless of intrinsic merit.

For the student of society the chief significance of this search for the basic cause of social stratification is its bearing on the possibility of reform. It may well be that the stability of the social order, to say nothing of the welfare and happiness of the average human being, would be improved if the conditions of life within the social system were more nearly uniform. The individual who would like to see some action taken toward this end will find that his attitude toward the possibility of reducing these contrasts through conscious social effort will rest upon his acceptance of one or the other of the theories of social stratification discussed in the preceding paragraphs.

Labor's share of the social income. The law of proportionality. Assuming that men continue to live and work in a society like our own, what prospect is there that labor, either in its entirety or within a given labor class, can increase its share of the social income? For society as a whole, improvement in real income can be due only to some general development affecting the physical productiveness of society's economic system. But it does not necessarily follow that an increase in the total income of society will be evenly spread over the several shares in distribution. It is quite possible that the forces which bring about increased general prosperity may so operate as either to increase or decrease the fraction of the whole which goes to labor, as well as the average real wage rate.

Our analysis of diminishing returns, applicable alike to rent, interest, and wages, brings to light a principle of universal truth with regard to the relative economic prosperity of the different income receivers in society. This principle may be formulated as follows: *Any social change which reduces the ratio which one factor or combination of factors bears to those coöperating with it will tend to raise the average income of that factor or combination of factors as compared with the others.* If population increases faster than land and capital, real wage rates tend to fall in proportion to rates of interest and rent. If capital accumulates faster than the increase in population and natural resources, rates of wages and rent tend to rise in proportion to the rate of interest. If the available supply of land and natural agents expands out of proportion to labor and capital, the unit rent income tends to shrink as compared with wages and interest. This generalization is called the *law of proportionality.*

Applying this principle to the economic welfare of the laboring population, it will be seen that average wages per worker, as a share of the whole income, will be increased by any development which expands the social income by increasing the quantity of productive factors other than labor.

The principle is applicable not only to changes in a factor of production taken as a whole in relation to the other factors, but also to changes of balance among the subgroups of a given factor. If within the labor supply, for example, managerial labor increases in quantity while other types of labor do not keep pace with it, the average real income of the one will tend to fall proportionately to that of the others.

This feature of our exchange economy has far-reaching effects which are not always favorable to general social welfare. It means that it is often to the immediate economic advantage of each separate labor group to restrict its membership and raise the value of its marginal product. When the workers in a given occupation are organized into trade unions and are able to prevent those who are not union members from engaging in the occupation, they are tempted to adopt restrictive practices which artificially limit their numbers and, in an equally artificial manner, increase the numbers in other lines of activity. This will raise their marginal productivity, make their product relatively scarce as compared with other products, and increase their average real wages.

Much the same result is obtained if the group in question adopts "make work" policies which set a limit to the output of the average man in a given time. In other words, the arrangements of our social order make it possible for specific labor groups to benefit themselves by working positive injury on society as a whole.

EXERCISES

1. The annual income realized by an automobile repair shop with different numbers of mechanics employed would be as follows:

Mechanics	Income
4	$45,200
5	51,000
6	55,200
7	58,100
8	60,000
9	61,200

How many mechanics would be employed at a wage rate of $150 per month?

2. Men's shoes are sold in a retail store at a gross profit or markup of $1.00 per pair. The salaries of salesmen are $160 per month. Assume that the following relationship exists between number of salesmen employed and monthly volume of sales:

Number of salesmen	Volume of sales
1	480 *pairs*
2	880
3	1,200
4	1,440
5	1,600
6	1,680

(a) Determine the marginal cost per unit of sales when different numbers of salesmen are employed.

(b) How many salesmen would be employed and what would be the volume of sales?

CHAPTER XXIII

Profits

Nature of profits. "Profits" is one of the commonplace expressions of everyday speech; yet it is rarely given a fixed or definite meaning. Like many other technical terms of economics, it requires careful definition in order to avoid confusion of thought. As used in economics, the term denotes a type of income distinct from the other types which have been considered in the preceding study of functional distribution; *i.e.*, a share of the social income separate from rent, wages, and interest. It is our problem now to identify and define this particular category of income and to consider the forces which account for its existence and determine its amount.

The accountant's profit statement. As a point of departure, let us place ourselves in the position of a typical businessman recording the results of his year's activities. Let us suppose that he is a retail coal-dealer; that he leases all the land and borrows all the capital employed in his business; and that, instead of managing the business himself, he employs a manager. The following account records in round numbers the income and outgo of the business for a certain year.

PROFIT AND LOSS STATEMENT, 1947

Net sales		$100,000
Cost of goods sold	$75,000	
Selling and administrative expense	20,000	
Total costs of operation		95,000
Net operating profit		$ 5,000
Other income (none)		
Other expense		
Rent	$ 1,000	
Interest	2,000	$ 3,000
Net profit		$ 2,000

This account will suffice to indicate the essential nature of profits. It will be noted that the accountant uses the term "profit," with appropriate qualifying terms, in at least two senses. The net results of

the business as an operating entity are shown by the item, "net operating profit." But against this we must recognize the fact that this business required that certain land and capital be devoted exclusively to its uses, thereby being withdrawn from other possible remunerative employment. The use of these factors could not have been obtained except by agreeing to pay their owners economic rent and interest respectively. And it is only after deducting these payments that we arrive at the true net earnings which go to the owner of the business as such. The accountant calls this "net profit," and this is true *profits* in the economic sense.

Implicit rent, interest, and wages of management. Let us now take another retail coal business, precisely like the one we have considered, except that the land and capital belong to the owner and the owner manages the business himself, adding the further assumption that the manager's salary in the first enterprise was $5,000. In this case, since the owner has no payments to make for interest, rent, or manager's salary, it might appear that his profit and loss statement should read thus:

Net sales		$100,000
Cost of goods sold	$75,000	
Selling and administrative expense	15,000	
Total costs of operation		90,000
Net operating profit		$ 10,000

In harmony with popular usage, which assumes that profits are what is left to the owner out of his business after deduction of payments for costs, the layman might regard this $10,000 as the "profits" of the year's business. But in our search for the true profit, we need to remember that the land site owned by this coal-dealer could have been rented, had he so chosen, to another entrepreneur for $1,000 a year, and that the capital funds, which he had temporarily tied up in buildings and other capital instruments (trucks, shovels, coal chutes, etc.) and a stock of coal, might have been invested so as to yield him interest to the amount of $2,000 per year. Let us also recognize that the coal-dealer himself, instead of devoting his time to this particular enterprise, might have hired himself out to another coal company at an annual salary, let us say, of $5,000. In view of these considerations the economist insists that this particular enterprise has not really brought in a profit of $10,000.

From this net operating profit of $10,000 there must be deducted the rent, interest, and owner's salary, all three of which could have been obtained by the owner had he not gone into the coal business for himself. Only the remainder, $2,000, can be said to have come to him on account of the business. This is the true profit. We thus reach the general conclusion that, in economic analysis and in all accurate accounting, before true profits emerge, due allowance must be made, not only for explicit costs (payments to nonowners), but also for the implicit rent, interest, and wages of management where these costs are not represented by payments to others.

Definition of profits. We have now arrived at a clear and exact concept of profits and are in position to formulate a definition. *Profits are the net income of a business, or the difference between the income and the costs, the costs including rent, interest, and wages of management, whether explicit or implicit.* We must note that profits are essentially variable and uncertain. In any given enterprise, they may be large one year and small the next. Costs may exactly equal income, in which case profits are zero, or may even exceed income, resulting in losses instead of profits. We shall find it an aid to our analysis to regard losses as negative profits, which makes it possible to say that profits are always obtained by subtracting costs from income, even though the former be the greater and the result therefore a negative quantity.

Profits a residual income. A rather common error comes from thinking of profits simply as one of the "shares" of the gross income of business, on a par with wages, rent, and interest. That this is an error appears as soon as we ask: Who pays profits to the businessman or the corporation? Of course the answer is that nobody pays profits to anyone.

Let us consider a man starting a grocery store. We may suppose that he borrows money with which to buy a building, furnishings, stock in trade, etc., and for working capital, agreeing to pay interest at a stated rate to the lender. Or we may assume that he uses land and buildings belonging to someone else, in which case he agrees to pay rent to the landlord. He hires salesmen and office help, agreeing in advance to pay wages to each at a certain rate.

As time goes on our storekeeper keeps up payments at the agreed rates to his hired help and to those from whom he has borrowed capital or rented land, buildings, or equipment. No one, however, is currently paying him for his own labor, or for the use of any cap-

ital or land of his own which may be devoted to the business, or for his special entrepreneurial services of planning and organizing.

The entrepreneur's reward comes only as the financial results of the business emerge. If after paying from the gross receipts the wages, interest, and rent due to others, he has something left over, that is his net income. If the gross receipts are not enough to pay these costs, he suffers loss. Profits (and losses) are the end result of the enterprise, a contingent reward, depending on the degree of success of the business.

Capitalists, landlords, and laborers are induced to perform their respective parts in production by the advance offer of interest, rent, or wages, at certain rates. For the entrepreneur, on the other hand, there is no such thing as a "normal rate" of profits, or even any "rate" of profits at all, except as a historical record or estimate of the ratio of past profits to capital invested or some other base. What leads the entrepreneur to perform his function is, strictly speaking, not profits but the chance of profits.

Business risks. The distinguishing feature of profits as a functional share in distribution is that they arise out of the uncertainties of modern business. Business risks may be roughly classified in two main groups: (1) those which are incidents in routine business operations, and (2) those which arise out of dynamic changes in the general structure of industry and society. The former concern individual enterprises within an industrial field; the latter exert their effects upon industries in their entirety.

Ordinary business risks, in so far as they become foreknown, lose their uncertainty and can be handled as costs. The development of the practice of insurance makes this clear. When it is discovered that a given risk is insurable, the losses which it entails are apportioned among those affected by it. Payment for insurance then becomes one of the normal and calculable costs of doing business and plays its part in determining market prices.

But not all of the routine risks of business are thus amenable to insurance, and in addition there are the more important uncertainties, generally uninsurable, which are attendant upon growth, change, or decay in industry or society at large—changes in industrial processes over short and long periods, changes in the nature and range of human wants, changes in the availability of natural productive agents.

The risk-bearing function. The vast majority of us take no thought about these matters. The work of anticipating future needs, discounting future hazards, making timely adjustments to forthcoming alterations in the conditions of economic life is left to a self-constituted body of men who possess the requisite foresight, courage, and energy. We may call these men the risk-bearers if we are careful to keep in mind the broad implications of the term; but their major social function is that of guiding and directing the course of economic development. Those who draw profits from the anticipation of future costs and income are performing a good deal broader function than mere risk-bearing, as that term is ordinarily used. It would be more illuminating of their real function in society to say that *they capitalize their prevision of the future.*

Consider, for example, the function of underwriting new ventures, which is discharged by the investment banks. When they believe that timely provision of capital equipment along novel lines promises lucrative returns in the future, these banks guarantee to provide the necessary supplies of present funds, partly out of their own resources, partly by inducing others to take a share in the venture. Through their efforts, capital instruments are constructed and assembled, managerial and manual labor is organized, the different parts of a productive establishment are brought into coöperation with each other. If the forecast is correct, this productive unit will be equipped and ready for service at the time it is needed. The bankers may then withdraw and surrender the actual ownership and management of the new enterprise to others. But they will have capitalized the opportunity for profit by retaining for themselves shares of stock in the enterprise, whose dividends absorb a large part of the pure profit which accrues.

The process we have just described is known in the business world as *promoting* new industries. Professional promoters, called into existence by the dynamic character of modern industry, devote their whole energy to forecasting the currents of change at work and to placing themselves in position to profit from the sale of goods and services by an excess of future income over the expenditures, entailed in costs of production. A new railroad is built across a stretch of sparsely settled prairie at a cost of scores of millions. Its promoters are predicting a thickening of population in the region, a shifting of the geographical location of industry, a new relation

between producing and consuming markets. If these changes occur, the promoters will have rendered society a service by providing capital instruments of immense worth in the new setting, and they will profit accordingly.

Another foresees that the need of a great nation for meat products cannot long be satisfied piecemeal from local sources of supply. He constructs a plant for the curing and preserving of meats, perfects the technical processes, elaborates the buying and selling mechanism, and profits with the coming of the conditions which he has foreseen. Illustrations of the promoter's activity could be multiplied indefinitely.

But we must not allow the glamour of "big business" or the spectacular risks borne by the great promoters to obscure the fact that the service of which we have been speaking and the profit which is its recompense are involved in the most commonplace and picayune of business operations. The little corner store which appears in the midst of a straggling suburban settlement signifies someone's faith in his foresight; it may be wiped out by a shifting of industries, or a change in transportation; it may flourish on an increasing need for the services of a middleman in the little community. Every independent entrepreneur who embarks on a new line of productive service, or adds to the existing facilities in the older lines, is to some extent relying on his prevision of future events. Success will bring him excess of income over and above the necessary payments for his labor of management; failure will deny him this excess of income and may sweep away both the reward for his labor and his entire investment of capital.

Actual profits and losses. Of the actual existence of profits as a type of social income, there is ample evidence. During the twenty-year period between 1901 and 1922, the United States Steel Corporation accumulated surpluses of over three quarters of a billion dollars in addition to dividends paid to stockholders. In one ten-year period, the Standard Oil Company of New Jersey paid annual dividends of 20 per cent to its common stockholders and accumulated several hundred millions out of earnings. Henry Ford began his manufacture of motor cars with an investment of $49,000, and within twenty years had reinvested earnings amounting to nearly half a billion dollars. These earnings were by no means all profits, since they must be assumed to include interest on the capital invested. But even though no precise segregation is possible, there

can be no doubt that very large profits were earned by these companies.

Such spectacular examples, however, should not be allowed to stand alone. The other side of the picture discloses an impressive record of losses, or negative profits.

The records show that every year, even in times of business prosperity, many thousand corporations have no net income out of which to pay dividends. The following table, compiled from reports of the United States Treasury Department, uses the term "net income" in the sense of income subject to federal tax. Corporations with no such net income obviously had negative profits in the years given.

CORPORATIONS WITH AND WITHOUT NET INCOME[1]

Year	With net income	Without net income
1909	20 *per cent*	80 *per cent*
1913	60	40
1917	66	34
1921	48	52
1925	59	41
1929	53	37
1931	34	55
1933	22	67
1939	39	52
1941	52	40

[1] Before 1926, the Treasury's figures for corporations without net income included inactive concerns. These were excluded after 1926; hence, after that date, the percentages do not add up to 100.

The years given in the table have been selected to include both good times and bad. The years 1917 and 1929, for example, were characterized by exceptional business prosperity; yet in those years about a third of the corporations failed to earn enough to pay interest on capital invested by the stockholders, to say nothing of pure profit. In times of business depression, for example, 1921 and 1933, the cases of corporations without net income are multiplied. In the latter year two-thirds of all corporations had no net income.

Negative profits are shown most strikingly when a business enterprise fails—*i.e.*, goes into bankruptcy. During the period 1900 to 1939, the number of failures averaged 16,000. In 1922 there were 23,676 failures, with total liabilities of 625 million dollars; in 1932 the number of failures was 31,822, and total liabilities 928 millions. In addition to the enterprises that go bankrupt, many others each

year are so discouraged by their inability to earn a profit that they wind up their affairs and withdraw from business. During the period 1900 to 1939, such annual withdrawals averaged 360,000 or about one-fifth of all business concerns. The record since 1939, both for withdrawals and failures, is given in the following table.

BUSINESS WITHDRAWALS AND FAILURES[1]

Year	Withdrawals	Failures
1939	349,000	15,000
1940	337,000	14,000
1941	329,000	12,000
1942	307,000	9,000
1943	296,000	3,000
1944	352,000	1,000
1945	255,000	1,000

That the risk element in business enterprise is not limited to especially hazardous businesses, but pervades the entire system, is shown by startling reverses of fortune which affect the entire industrial world. As an index to this situation, we may take the aggregate net income accruing to all corporations when losses have been subtracted. In 1918 this aggregate net income was 8½ billion dollars; in 1921, barely more than 1 billion. Rising to 12 billions in 1929, aggregate net income fell to 4⅔ billions in 1930, and in the next three years there were net losses, reaching 3,829 billions in 1932. Aggregate net income was 7⅘ billions in 1937, 4 billions in 1938, a little over 7 billions in 1939, and 23 billions in 1942 (this year showing the effect of war activity). We have to remember that these figures do not represent what the corporations had at their disposal, since federal taxes have still to come out. In 1942, for example, the federal taxes took over half, *i.e.,* 12¼ billions.

The historical "rate" of profits. A statement issued in 1945 by the United States Treasury Department makes it possible to relate profits to the capital invested in business.

In the year 1937, the combined net profit of 416,902 corporations which submitted balance sheets with their tax returns was 7,777 million dollars. This was at the rate of 5.8 per cent on the stockholders' investment (net worth at end of year). In the three years, 1931 to 1933, as we have seen, the combined result of the business activities of all these corporations was a net loss. Averaging the

[1] National Industrial Conference Board, *The Economic Almanac for 1946–47*. New York; the Board, 1946, p. 105.

nine years for which these figures are available, 1931–9 inclusive, we have an annual net profit for these corporations (averaging some 405,000 in number) of only a little over 2½ per cent on the stockholders' investment.

We must always remember, however, that these figures and the others we have been using deal with net income and not with profits as such. Even when a corporation ends the year with a net income, this fact alone does not demonstrate the existence of pure profits after normal interest on invested capital has been allowed. Thus if we recognize that 2½ per cent is at best only a moderate rate of interest, it appears that during the period 1931 to 1939 the stockholders of American business corporations as a whole scarcely recovered interest on their investment, with little or nothing to show in the way of profits.

Profits represent also only a small part of total corporation receipts. Official reports of the United States Bureau of Internal Revenue show that, for the twenty-six years, 1916 to 1941, the total net income less deficit of all the corporations reporting to the Commissioner of Internal Revenue averaged only 4.5 per cent of their combined gross income. Only in the war years, 1916 to 1919, did the percentage ever go as high as 6 per cent. Gross income is not precisely the same as income from sale of goods and services, but it is close enough to warrant the conclusion that over the years not more than 6 cents out of the consumers' average dollar goes to profits.

Profits tend to be short-lived. Pure profit, when it does arise, tends to be a temporary income. The forces of competition work incessantly toward a readjustment of prices which will wipe out the margin of income above costs. Chance gains resulting from lucky ventures or the caprice of fortune are by nature limited in duration to the continuance of the special events which have produced them. Special devices or methods of management, by which individual businessmen avoid loss from routine business hazards, become known to the trade in general and lead to an adjustment of prices. Even the larger flows of profits resulting from far-reaching changes in industry are of limited duration. A new invention will bestow profits on its first possessor by reducing his costs below those of his competitors. But when it becomes the common property of the trade, all costs will fall and the maladjustment will disappear.

The long-run effect of improvements in machine equipment and

in processes of manufacture is not to expand the profit income, but to increase the real incomes of all groups in society. Even the forerunners in new fields of industry, who are the chief profit makers in the business world, are not long exempt from competition. Those firstcomers may withdraw with large amounts of capitalized profits before the leveling tendency of competition has brought prices and costs into adjustment, but the industry as a whole soon ceases to be a source of profit to those who exploit it.

Profits not a price determinant. Pure profit is not one of the price-determining costs of production. This income may not exist as a positive quantity for industry as a whole over a term of years. The service of the profit maker is induced by the chance of receiving the reward, not by the certainty of it. In any hazardous undertaking, it is the anticipation of the reward which induces the efforts of the competitors. In athletic contests, a winner's prize need be given to but one of the contestants. It is doubtless true that the service for which pure profit is the reward would not be rendered if none of the entrepreneurs were able to make a profit. But by holding forth chances of gain to individuals, society is supplied with the risk-bearing service of the entrepreneurs as a class without guaranteeing a reward for this service in any particular case.

Recalling again that we speak only of competitive industry, the following statement, analogous to the case of economic rent, can be made: Prices are not high because profits exist; but profits exist because for other reasons sale prices are sometimes higher than costs.

Criticism of the profit system. Yet it is probably true that no feature of the capitalistic order of society has been subjected to more aggressive criticism from the advocates of social reconstruction than that of the profit incentive. The more superficial of these criticisms are aimed at the size of the profit income. It is carelessly assumed that the profit income, both as a whole and in individual cases, is so large as to create great and unjust inequalities in the economic conditions of the different social classes. That this income is large in certain individual cases we have had occasion to note. But the customary indictment of the profit income on the score of its size fails to take adequate account of the burden of losses borne by the same social class which is the recipient of profit. As we have seen, the total of profits is a very moderate sum in relation to total investment or total sales.

The more weighty criticisms of the profit system, however, are concerned, not with the size of the income, but with the prerogatives and powers of the profit-getting class. The receipt of profits is associated with ownership and control. This ownership draws a line of division between the profit-getter and the worker, which places the latter under the direction of the former, and without giving him a share in control, renders him liable to injury through mistakes in judgment on the part of the owner or his representatives.

This allocation of power, however, is inevitable in the present system. The risk-bearer must be the commanding officer of business enterprise, else he is powerless to perform his function. Foresight divorced from power can have no practical influence over the adjustment of the industrial structure to changing conditions.

In connection with our study of the other shares of distribution—rent, interest, and wages—we found that certain features of these institutions of our society gave rise to organized movements for change. Criticisms of the profit system provide additional stimulus to these movements. In a later chapter we shall examine in detail the programs of reconstruction to which these criticisms have given rise.

EXERCISES

1. Mr. X formerly was employed as department head of a manufacturing enterprise at a salary of $6,000 per year. He had $20,000 of savings invested in bonds which yielded him 5 per cent interest. He resigned his position and set up an enterprise of his own, investing his savings in the new enterprise. At the end of the first year, his net income from the business amounted to $6,200.
 (a) What were his pure profits? Explain.
 (b) If his net income continued at this rate for ten years, would it be worthwhile for him to continue in this enterprise? Explain.
 (c) If this rate of return were typical of earnings of the industry as a whole, what would be the trend of future development in the industry? Why?
2. Suppose that Mr. X in the following year receives a net income from the enterprise of $10,000.
 (a) What now are his pure profits? Explain.
 (b) If this rate of return were typical of earnings of the industry as a whole, what would be the trend of future development in the industry? Why?
3. A new type of machine costs $1,000 and yields an annual net return of $120 on the investment. If the prevailing rate of interest is 5 per cent, what amount of pure profits will result from the use of the machine? Under conditions of competition, will the income from the new machine tend ultimately to be equalized with the income from other machinery? How?

CHAPTER XXIV

Government. Public Income. Taxation

Government's place in the economy. Far back in the obscure beginnings of human existence—whatever and wherever they may have been—there was the beginning of government. The human race has always been socially inclined, living its life not as isolated individuals, but in organized groups: families, tribes, clans, towns, and nations.

For such organized existence, there had to be authority. The philosophical concept of anarchy—no government—is in practice nowhere to be found, nor was it ever to be found so far as history records. So intimately is government related to every branch of the economic life of a people, that the student of economics is compelled to take cognizance of government as an essential feature of any economic system.

At different epochs in the history of mankind, various forms of government have appeared, passed across the stage, and made their exit. At the present time, many kinds of government are to be found in different parts of the world, from the rule of the tribal chief or medicine man or the soviet rule of Russia to the representative government of the United States.

In harmony with the purpose of this book, our interest is chiefly in that kind of government which goes with a free economy. Such an economy requires on the one hand that the essential functions of government be provided, and at the same time that the government's powers and actions be held within strict limits.

For example, the American nation was established following a successful revolution against governmental interference with the economic aspirations and activities of a freedom-loving people. During the Revolution, and for a few years thereafter under the Articles of Confederation, the American people suffered because their national government lacked sufficient power to perform its essential functions. They soon came to realize that their progress in the production of wealth by exploiting their fabulous heritage of

natural resources was being retarded for lack of the coöperation of an adequate national government. So, in 1789, they adopted the Constitution of the United States, which made provision for a national government with powers adequate for the performance of its essential economic functions.

But these early Americans were at the same time very fearful of too much government, and a large part of the Constitution is devoted to restrictions upon the powers of their new government and "checks and balances" to prevent the assumption of excessive authority. Later the Bill of Rights was added, consisting chiefly of a series of limitations upon the national government in the interest of personal freedom.

American political organization. Whereas in most countries the national government is the sole fountainhead of sovereign power, the United States is a federation, in which sovereignty is divided between the national government and the several states. We are accustomed to speak of three grades of government: (1) national or federal, (2) state, and (3) local.

The federal government was formed in 1789, after the people had adopted the Constitution of the United States. Before that there had been thirteen states, which had just won their independence from Great Britain and which were therefore separate sovereign powers. These independent states voluntarily surrendered some of their sovereignty in order to form the national government, which has only such powers as are to be found either granted to it in the Constitution of the United States or implied as corollaries of the powers so granted.

The states, increased now to forty-eight in number, are the original and residual possessors of sovereignty, having all sovereign powers except such as have been granted to the national government. The third grade of government has no sovereignty, but consists of administrative units, counties, towns or townships, cities, boroughs, school districts, road districts, fire districts, irrigation districts, etc. These units, created by the states, hold only powers which have been granted them by their respective states and which may be taken away at the state's will.[1]

[1] The term "government" is used in this book to include all grades of government of any particular community: for example, in the United States of America, the federal government, the state governments, and the governments of counties, towns or townships, cities, school districts, and all other political

The foregoing summary of the political organization of the United States very probably contains nothing that is new to the reader. It is merely a reminder of facts which are vital to the subject we are about to study.

The functions of government. In the current controversies over the character of government and the scope of its activities, there is some danger that we may overlook the ordinary everyday operations of government. It may be well, therefore, to summarize briefly here what we may conveniently call the primary functions of government. These include (1) those of a protective nature: defense against outside foes, requiring army, navy, etc.; justice and security, afforded by police service, courts, and penal institutions; regulation of private industry, in particular monopolistic industries, like the railroads, and those which are so complicated as to be only imperfectly understood by the public, such as the banks and insurance companies; protection against disease and accident; maintenance of moral standards; and protection against fire, storms at sea, and certain other forces of nature. Then (2) there are the developmental functions: education; religion; facilities for private industry, especially highways and bridges, harbors, the monetary system, information services, etc.; and development and conservation of the nation's natural resources.

In addition to the foregoing, which are peculiarly governmental functions in modern times, government may enter the industrial field, where private enterprise generally holds sway, and we may then have a considerable variety of government industries.

Government income and costs. To meet the heavy expenditures necessitated by the performance of its functions, the modern government must have an equally large income. In modern times there are just four important sources from which government income is to be obtained, and these now demand our study. They are (1) industrial earnings (prices), (2) taxes, (3) fees and assessments, and (4) loans.

The first three of these are called *revenues,* meaning receipts, in money or its equivalent, which increase the government's assets without increasing its liabilities, or decrease the liabilities without decreasing the assets. They are thus distinguished from *loans,* which

jurisdictions. It is rather common in popular speaking and writing to refer to the United States federal government as "the government" in distinction from the states and the local jurisdictions. The term is not so used in this book.

are receipts that cause an equal increase in the government's liabilities.

Revenues correspond to *costs,* which are government payments, in money or its equivalent, which decrease the government's assets without decreasing its liabilities.

When revenues equal costs, the government is paying its way or balancing its budget. When revenues are less than costs, there is a *deficit,* and sooner or later resort must be had to borrowing. When revenues exceed costs, there is a *surplus.*

Financing government industry. When a government engages in a business enterprise, such as the United States Post Office or the water supply system of a city, it sells its product (goods or service) at a definite price to the consuming public, much as any private business concern does. The citizen may take the offered service or not as he wishes, and if he does not take it he makes no payment. Payment is voluntary and in proportion to the amount of the service received. For example, the government offers, through the post office, to transmit letters and parcels. If the citizen chooses to take advantage of this service he pays the stated price; otherwise he pays nothing to the post office. In these respects, government industry is similar to private business.

Financing the primary functions of government: General rule. If now we transfer our attention from the industrial undertakings of government to its primary functions, we shall find certain fundamental differences in their financing. The primary functions are undertaken by the government in order to benefit the people as a whole, without the purpose of giving benefit to particular individuals. For example, the state and local governments of the United States provide public schools and universities at great expense, not primarily in order to benefit those persons who want an education for themselves or their children, but because it is commonly believed that a prosperous and enduring democracy is not possible unless there is widespread education among the people. Education is provided therefore to benefit the whole people, not for the sake of those particular individuals who attend the public schools.

No attempt therefore is made to pay the cost of education by charging a price and so collecting from each citizen in proportion to the benefit he has received. The government does not sell education as it sells postal service; education is offered freely to all. The

burden of furnishing the money necessary to pay the cost of public education is distributed among the people according to certain rules, which we shall presently study; but there is no necessary relation between the amount of any person's contribution and the service received by him, nor may any person avoid contributing by declining to avail himself of the government's service. Contribution is compulsory.

Education illustrates the general rule followed in financing the primary functions. No attempt is ordinarily made to obtain, in connection with each function, a revenue earmarked to meet the expenditures of that particular function.

Taxes. The foregoing analysis of the purpose of the government's primary functions and the financial contrast between such functions and the government's industrial undertakings has introduced the principal characteristics of taxation: namely, (1) compulsion, (2) devotion to the general welfare; and (3) apportionment without reference to individual benefit. It will be found helpful to have a precise definition of a tax, and the following one will serve our purpose: *A tax is a compulsory payment by a person to the government, destined to defray the cost of government services performed for the common benefit, and paid without reference to individual benefit.* It is upon taxes that modern governments principally rely for obtaining the money necessary to defray the cost of performing their primary functions, and taxes make up the bulk of the revenue of most modern governments.

Fees. The primary functions of government are performed, as has been shown, for the common benefit. Nevertheless, it often happens that, in performing its primary functions, the government incidentally gives special benefits to particular individuals.

Consider, for example, the granting of patents for inventions. The purpose of this service of government is not to give a special benefit to the inventor; if such were its purpose it would be a case of favoritism, repugnant to the spirit at least of democratic government. The real purpose of the patent is to promote the public welfare by encouraging the discovery and disclosure of useful new devices and by offering inducement to capitalists and entrepreneurs to develop and market them. It is for this reason that the inventor is given for a time a privileged position of monopoly with respect to his discovery. Nevertheless, this is a material benefit conferred upon the inventor, and for that reason the government imposes

upon him a special charge, justified by the particular benefit which he is presumed to have received. Such a payment is called a *fee.*

Fees arise also from somewhat different circumstances. There are certain occupations which cannot be left to the free and unrestricted operation of all persons without danger to the public interest. Such are the businesses of banking, insurance, taxicab operation, peddling, etc. Now the performance of this regulatory function necessitates government expenditure, and the person who chooses to enter such an occupation is thereby causing expense to the government. The government therefore charges such person a fee, covering part or all of the cost imposed upon it by his engaging in the regulated occupation. Fees are thus imposed upon those who engage in any activity over which some government regulation has been found necessary. One cannot drive an automobile, or own a dog, or enter the marriage state, without a government license; and in each case he is charged a fee to cover the cost which his act has imposed upon the government.

We now have in mind the nature of the fee, which may be defined in precise terms as follows: *A fee is a payment by a person to the government on account either of a special benefit received from the government or a special cost imposed upon the government in connection with a government service performed for the common benefit.*

Special assessments. Suppose the city authorities decide to put a sewer in a certain street. The consequent rise in value of land fronting that street will bring material advantage to the owners of such land. The same rise in land values follows the introduction of paving, street curbs, sidewalks, and other improvements. The government undertakes a new sewer, not from any desire to benefit particularly the owners of the land, but because the interests of all the people of the city will be promoted by better sewage disposal on that street. Here then is a perfect occasion for a fee and, in America at least, a fee is usually charged, which is none the less a fee in spite of its common designation as a *special assessment.* This is the definition: A *special assessment is a particular kind of fee, imposed when the special benefit consists in an enhancement of the value of land.* The amount of the special assessment may be anything up to the total addition to the value of the land caused by the government improvement. The special assessment furnishes a most interesting example of the fee, since there is here usually an attempt

to relate (more closely than for fees in general) the amount of the payment to the value of the special benefit conferred.

Taxation, general concepts: Technical definitions. Now it is time to acquire precise notions of certain technical concepts necessary to the investigation of the subject of taxation.

The *object* or *base* of a tax is the thing or the fact upon which the tax is based. Land, houses, and other articles of wealth are objects of a property tax. Each person's taxable income is the object of the income tax. When a rate of tax is imposed, the amount of such tax is found by multiplying the numerical statement of the object or base by the *rate.*

The object of a tax is to be distinguished from the *source,* which is the wealth or income from which the tax is paid. Much confusion can result from failure always to keep this distinction clear.

We have also to distinguish between the *immediate payer,* the one who actually makes payment to the tax collector, and the *subject* of a tax, who is the one who ultimately bears the burden.

These notions can all be illustrated by reference to the United States tax on cigarettes. This tax is imposed by the federal government, at the rate of $3.50 per 1,000 cigarettes. On each packet of 20 cigarettes there is a tax of 7 cents. The object or base is the packet of cigarettes. This tax is collected from the manufacturer of the cigarettes, and the fact of its collection is evidenced by the stamp attached to the packet. The manufacturer then is the immediate payer. But he does not bear the main burden, since he adds the tax to his selling price; if a packet sells for 20 cents, 7 cents of this price is tax. The consumer bears the burden; he is the subject. Finally the source is whatever fund the buyer draws upon for his cigarette purchase.

The foregoing paragraph calls attention to the fact that there is often a *shifting* of the burden of a tax from the immediate payer to some other subject. The final resting place of the burden of a tax, whether it is shifted or not, is called its *incidence.*

Taxes are usually put into two classes according as the burden is or is not normally shifted. *Direct taxes* are those whose burden remains on the whole upon those from whom the tax is originally collected. *Indirect taxes* are those whose burden is normally shifted. Since the possibility of shifting may be present only partially and in various degrees, it is not always possible to draw a sharp exclusive line between the two classes of taxes. However the char-

acteristics of the more important kinds of taxes are so clearly marked that there is usually little difficulty in placing any given tax in its proper class.

This distinction, be it noted, is an economic one. It does not agree precisely with the legal definition of a direct tax, as laid down by the Supreme Court of the United States. There is no occasion here for dispute. We have simply to deal with two different concepts, the economic and the legal, and must be on our guard only against the danger of confusion from the use of the same technical terms in two different senses.

Relation of tax rate to base. Great importance attaches to whether the relation between the amount of a tax payment and the base is one of simple proportion or is something different.

Suppose that A has a net income valued at $2,000, while B's is worth $4,000, and C's, $40,000. Does this indicate that a tax on incomes would require B to pay twice as much as A, and C ten times as much as B and twenty times as much as A? If so we have *proportional taxation,* to be accomplished by a uniform rate. Suppose the rate is 4 per cent. Then we have this result, in which the tax payments are in direct proportion to the respective net incomes:

Person	Net income	Tax rate	Amount of tax
A	$ 2,000	4 *per cent*	$ 80
B	4,000	4	160
C	40,000	4	1,600

If instead of making the tax payments directly proportional to the base, they are made to increase at a rate faster than the increase in the base, we have *progressive taxation.* For example, suppose a scheme of income tax rates such as this:

On the first $1,000 of the net income		1 *per cent*
On the part of the net income:		
over $ 1,000, up to and including	$ 2,000	2 *per cent*
over 2,000, up to and including	3,000	3
over 3,000, up to and including	5,000	4
over 5,000, up to and including	10,000	5
over 10,000, up to and including	20,000	6
over 20,000, up to and including	40,000	7
over 40,000, up to and including	60,000	8
over 60,000, up to and including	80,000	9
over 80,000, up to and including	100,000	10
over 100,000		11

Now A, with a net income of $2,000, will pay a tax of 1 per cent of his first $1,000 plus 2 per cent of his second $1,000; *i.e.*, $10 + $20 = $30. Similarly B's tax would be calculated thus:

1% of the first	$1,000	$ 10
2% of the second	1,000	20
3% of the third	1,000	30
4% of the fourth	1,000	40
Total tax on	$4,000	$100

C's tax would amount to $2,390.

These results may be tabulated thus:

Person	Net income	Tax rate	Amount of tax	Ratio of tax to net income
A	$ 2,000	1–2 *per cent*	$ 30	1.5 *per cent*
B	4,000	1–4	100	2.5
C	40,000	1–7	2,390	5.975

Under such a set of rates, B's tax would be more than three times that of A, though his income is only double A's. C's income is twenty times A's, but his tax is practically eighty times as great. This is the result of progressive taxation and is accomplished, be it noted, by giving up the uniform rate in favor of a rate which increases with increases in the base.

It is possible to apply the opposite rule, by fixing a scale of rates which decreases as the base increases. This is called *regressive taxation*. This plan of taxation never appears consciously embodied in modern tax legislation.

When there is exemption of a certain amount of the base of a tax, there is produced an effect of mild progression, even though the tax is technically proportional. Suppose, in connection with a 2 per cent property tax, there is an exemption of $1,000 of otherwise taxable property. A has taxable property to the value of $2,000; after deducting the amount exempt, he pays 2 per cent on the balance of $1,000, or a tax of $20. B, with $10,000, pays 2 per cent on $9,000, or $180. Though the rate on the taxable base is 2 per cent for both men, on the total amount of property possessed A's tax is only 1 per cent, while B's is 1.8 per cent. As the taxpayer's property increases, the rate approaches, but never reaches, the limit of 2 per cent. Such a condition is sometimes called *degressive* taxation.

Another concept of progression. In discussions of the problem

of tax apportionment, use is often made of somewhat different concepts of progressive and regressive taxation. According to such concepts, taxation is considered progressive when the amount paid by the individual increases at a rate faster than the increase in some assumed general or composite measure of economic power; roughly "the tax bears more heavily upon the rich than upon the poor."

In this sense the progressive income tax (even a proportional income tax with its ordinary exceptions, credits, etc.) is progressive. Consumption taxes, on the other hand, are generally considered regressive (even though they are imposed at uniform rates) because, in the first place, they are usually levied upon articles of wide general consumption which call for a greater share of the poor man's expenditure than of the rich man's, and secondly, they impose no burden on that part of income which is saved, a feature of which the poor can take only slight advantage.

This concept of progression, though lacking in precision, is often useful, especially when we seek to characterize a tax system as a whole. For this purpose it is not enough to know that particular taxes are proportional or progressive in the technical sense. (Technically, there are no regressive taxes.) What we are here concerned with is the general nature of the whole tax system in its relation to people on different income levels. In any tax system (using our second concept of progression), there may be, side by side, proportional taxes, progressive taxes, and some that are regressive. These taxes tend more or less to counteract each other.

If the whole tax system bears with equal severity upon all classes, we call it proportional. If as we go up the economic scale, the tax burden increases more than proportionately, we call the tax system progressive; the reverse of the latter system we call regressive. It is in this sense that these terms are generally used in popular discussion of tax problems. A precise quantitative measure of the net result, in any particular case, determining whether the whole tax system is proportional, progressive, or regressive, is extremely difficult, if not impossible. However it may be possible thus to characterize a tax system in a broad general way.

Principal forms of taxation. Among the various taxes employed by modern government, the following are the principal forms: property tax, income tax, consumption taxes (including internal excises and the customs tariff), death taxes, business taxes, and poll taxes.

The property tax: Evolution. The property tax is among the oldest of taxes; history shows us numerous examples in ancient and medieval times. At a time when there was not the bewildering variety of wealth and property that exists today, the early property taxes were based chiefly upon land and buildings. Later attempts to include various other forms of wealth and property rights, as they gradually evolved, proved unsuccessful, and the property tax in most European countries long ago returned essentially to its original base—real estate.

In America, on the contrary, where the property tax has always played an important role, the attempt is still generally made to include in its base most, or at least very many, of the complex modern forms of wealth and property rights. When thus inclusive, the tax is known as the *general property tax.*

The American property tax: General description. The property tax is prescribed by state law in each of the forty-eight states and the District of Columbia. Although no two states have exactly the same system, the systems are sufficiently alike to enable us to draw up a composite picture of the typical structure and operation of the property tax.

The first step is the *assessment; i.e.,* the discovery and valuation of all the taxable property in the jurisdiction of the taxing government as of a certain date each year. This function is performed generally by officers of the counties or towns acting under the state law. Assessment includes opportunity for appeal and revision in the case of taxpayers dissatisfied with the original assessment. Thereafter there is a process of *equalization,* performed by a county board where assessment is by the towns or, in other cases, by a state board. The purpose of this is to insure that the assessment in the several towns and counties has been uniformly done; *i.e.,* at full value or at the same percentage of full value. The assessment, as thus equalized, furnishes the official tax list or "grand list."

In most states revenue is obtained from the property tax by the state, the counties, the towns or townships, the school districts, and often by various other local taxing districts. Each such political body determines the sum which it will require from the property tax. By the process of dividing this sum by the total value of all taxable property within its own jurisdiction, it determines its "tax rate." This process is called *apportionment* of the tax. The tax due from any particular taxpayer is determined by multiplying the

assessed value of his property by a tax rate which is the sum of the tax rates of all the taxing jurisdictions to which his property is subject.

Position of the property tax. The property tax is an important source of revenue of the states and the principal reliance of the local governmental units of the United States, as is clearly shown in the accompanying table.

IMPORTANCE OF PROPERTY TAXES IN AMERICAN TAX SYSTEMS, 1942[1]

Governmental Units	TOTAL REVENUE RECEIPTS (In millions of dollars)	PROPERTY TAX RECEIPTS	
		Amount (In millions of dollars)	Per cent of total
State governments	6,114	274	4
Counties	1,635	897	55
Cities	3,123	2,026	65
Towns and Townships	371	257	69
School districts	1,798	1,078	60
Special Districts	148	90	60
Total	13,137	4,593	35

[1] Bureau of the Census, *Governmental Finances in the United States, 1942.* Washington: Government Printing Office, 1945.

In 1942 the 4,593 million dollars produced by property taxes for state and local governments represented about one sixth of the total revenues of all governments in the United States. This sum was exceeded only by the yield in 1942 of all sales and gross receipts taxes including customs.

Assessment: The problem. The principal problems of the property tax center about the assessment. The assessor is required (1) to discover all taxable property and (2) to determine its value. The first process, easy for real estate (land and buildings), heavy machinery, furniture, stock in trade, and similar forms of wealth, becomes difficult in case of such tangible personal property as watches and jewelry. Even such wealth as household furniture, books, pictures, clothing, etc., can be discovered only by a personal search of the taxpayer's home, a duty to which the ordinary assessor is loath to resort in the face of the natural repugnance of the taxpayer. When we come to intangible property—stocks, bonds, bank deposits, etc.—discovery becomes a practical impossibility.

Valuation presents even more difficulty. The typical local assessor, generally without expert knowledge of values, is required by law to appraise, within a period of a few weeks, all of the wealth in his town or county, including farms, factories of various kinds, complicated machinery, raw materials, "goods in process" and finished products of manufacture, stock in trade of the stores of utmost variety, and household effects of all sorts. In this task he will obviously not receive much aid from the taxpayers themselves, and he will always be under pressure to make his valuations low.

Expert assessment. In many of the larger cities and some small cities, a staff of trained assessors, with the necessary clerical and other assistance, is built up. The work is so arranged as to continue through the year, permitting of long-term or permanent employment, and giving to the assessor a steady "career job," with adequate salary. Accurate maps of the entire region are made and skillfully used. Complete descriptive data regarding land and buildings are prepared and recorded. Evidence regarding all sales is gathered. Scientific rules for determining values with respect to location and various other standards are used. Even the airplane is pressed into service for the quick mapping of extended areas in the suburbs or country districts. In many an American city where such conditions prevail, a high degree of efficiency, at least in the assessment of land and buildings, has been attained.

Some of the states have obtained a certain degree of centralization of the work of assessment in the office of a state tax commissioner or board, which has enabled them to secure full-time services of engineers and other experts to aid in the work of local assessment.

The ordinary town or county assessment. But in spite of assistance now and then from the state tax commission, assessment is still essentially a local function, and the ordinary town or county or small city will seldom have anything like an adequate assessment organization or procedure. There is not enough work to warrant a full-time job with attractive salary. Expert qualifications cannot be, and are not, required of the candidate for the assessor's job. There is lacking both the equipment and the personnel requisite for expert assessment.

Under such circumstances as these, we need not be greatly surprised at the result. The assessor generally does about the best he can. In the first place he has last year's lists, of which he takes full advantage; copying last year's figures will suffice for the majority of

cases where he has reason to believe there has been no material change. He will travel over his district, making brief inspections of the outside of new buildings, casting cursory glances at the merchants' shelves, asking a few questions, picking up what gossip he can as to the worldly goods of his fellow townsmen. Using such information as this, reenforced by a good deal of often shrewd guesswork, influenced almost always by the popular satisfaction which will follow when everybody feels that he has been assessed under the true value, influenced now and then no doubt by the desire to be kind to a friend and to stand well with those of political and social influence, seldom downright corrupt, the assessor manages somehow to check up on the lists handed in by taxpayers, to make out the lists of those who fail to file them, and finally to turn in a tax list which will not cause too many appeals to the board of relief, nor be too much censured by the county or state board of equalization, nor cause so much local dissatisfaction as to endanger his reelection.

Practical results. Under these circumstances it is not surprising that the accomplishment of the property tax falls far short of the ideal of the law. United States Census statistics show that the total assessed value of real estate is not more than half the true value legally taxable and that less than one-fifth of the true value of taxable personal property is assessed. These totals, moreover, give no picture of the great inequality among individual properties, which are assessed all the way from zero (where the property is not discovered) to more than 100 per cent of the value. As to intangible property, the breakdown in most parts of the country is nearly complete.

As certain properties escape, through failure to be listed or through undervaluation, the tax rates must be proportionally higher, and the burden is increased on other properties. On properties whose obvious face values do not permit of undervaluation, existing high tax rates (30 mills, 40 mills, or even as high as 80 mills) may be quite confiscatory and virtually force the taxpayer to concealment. Prevailing undervaluation moreover results in essential denial of the legal remedy for unjust assessment through appeal to the courts, since the taxpayer who is assessed at a higher ratio than his neighbors may still be assessed at less than true value and so less than the law requires. The attempt to correct wholesale undervaluation by the process of equalization generally fails, since county and state boards of equalization are rarely equipped to

make an accurate check of the original assessments in the towns or counties.

Confusing wealth and property. When a property right is expressed by a document, such as a stock certificate, a bond, or a mortgage, the property tax commonly includes the document as taxable property. This is *double taxation,* since it represents the taxing by the same government of the same object twice during each tax period. The present attitude of American state tax laws on this matter may be roughly summarized as follows: (1) shares of stock of domestic corporations are taxable to resident owners in somewhat less than half the states; shares of foreign corporations are similarly taxable in somewhat more than half the states; (2) bonds of corporations, whether domestic or foreign, are taxable to resident owners in about three-fourths of the states; (3) mortgages are generally exempt, almost always when secured by property within the state.[1]

Recognition of the peculiar position of intangible property, which has been pointed out, has led a number of states to provide for the taxation of such property at rates much lower than those applied to tangible wealth. It has been the hope that the taxpayers, when granted a reasonable rate, would voluntarily list their intangibles, and the experience of several states has shown some increase in the amount of such property assessed following a reduction in the rate. There is no evidence, however, to show that any state has by this expedient succeeded in assessing the whole, or even the greater part, of the intangible property subject to its tax laws.

Taxation of corporations. The taxation of corporations presents several special problems. In the first place, there is the complex problem (to which reference has been made) of double taxation resulting from taxing both the corporations on their property and the owners on their stocks and bonds. Second, it is the corporations which are the principal owners of those complicated categories of property whose valuation proves too much for the ordinary town or county assessors. In the third place, the wealth of many great corporations is located in more than one taxing district, being spread over several towns or counties or even states. The property cannot be assessed piecemeal, and no local assessor has authority to assess

[1] Cf. National Industrial Conference Board, Inc., *State and Local Taxation of Property.* New York: the Board, 1930.

the whole. This situation is peculiarly evident in the case of the railroads, telephone and telegraph companies, and other public service companies.

These problems have been met in many states by taking the assessment from the local bodies and entrusting it to a state commission. In addition special methods are frequently applied, very generally to the public utility corporations and the banks and less commonly to the general business corporations.[1]

Reform of the property tax. The first step toward a more efficient property tax is the abandonment of all taxation of intangible personal property. This is indicated both by the theoretical nature of intangible property rights and by the complete and inevitable break-down in administration. This step has been recommended by numerous tax authorities, including a committee of the National Tax Association, and it has been already accomplished by some ten states.

Next in order is the abandonment of the attempt to tax all those classes of tangible personality of which experience has shown the practical impossibility of a real assessment. This would include household furniture, private libraries, musical instruments, clothing and jewelry, and similar objects of essentially personal or domestic character. We may as well admit that these classes of property never have been and never will be properly assessed, and give up attempting the impossible.

A special problem is presented by machinery and goods and materials of manufacturers and merchants. We have the choice of either giving up the taxation of such goods, as has been done by New York, or consigning their assessment to the state tax commission, with authority to use, when advantageous, specified rules of assessment which do not necessarily arrive at market value. This second remedy has been recommended by a special tax commission in Connecticut.[2]

These changes as to personal property would leave for assessment by the local assessors only real estate and a few forms of bulky, easily discovered, and easily valued tangible personal property. Since the classes of property which it is proposed to exempt are the

[1] See later discussion of business taxes, pp. 478-79.

[2] *Report of the Connecticut Temporary Commission to Study the Tax Laws of the State and to Make Recommendations concerning Their Revision.* Hartford: State of Connecticut, 1934, Chapter x.

ones which now generally escape, the loss of revenue from this reform would not be great.

There are, as we have seen, certain cities whose assessment of lands and buildings has attained a high degree of efficiency. The organization requisite to produce this result is possible in any town or county where the finances are on a large enough scale to warrant the expense. For the smaller and less wealthy towns and counties, a consolidation of the work of assessment in the hands of larger units or groups of units is possible. There is no reason why any state should not provide assessment machinery adequate to accomplish its purpose; *i.e.*, the equitable distribution of the cost of government.

The property tax, with all its defects, is still the backbone of the American tax system. During the years of depression that followed 1929, it proved the anchor to windward that saved American public finance from a vastly greater shipwreck than it actually suffered. Its relative contribution, as compared with other taxes, increased materially during the depression. Fortunately the defects of the property tax are not incapable of remedy along lines which have been pointed out.

The income tax: Historical background. In contrast with the property tax, the income tax is a modern tax. The first real income tax was introduced in England in the year 1798. It was intended as only a temporary measure, though it actually lasted till 1816. Restored in 1842, the income tax has since then been a permanent part of the British revenue system. Italy, Germany, and many other nations introduced the income tax at various dates during the latter half of the nineteenth century.

In America the states made no use of this tax before 1840, and such state income taxes as were introduced thereafter in the nineteenth century were abortive and of no permanent significance. The federal government first made use of the income tax at the time of the Civil War. This was a war measure; its yield declined rapidly after the war, and it was repealed in 1872. In the year 1894, the national government made its second excursion into the field of income taxation; this act was declared unconstitutional before it went into effect.

It was not until 1913, following an amendment of the federal Constitution, that the income tax became a permanent part of the tax system of the United States Government. About this time, state

income taxation (beginning with the Wisconsin law of 1911) also acquired a fresh lease on life. Today the income tax is the principal tax of the federal government and a source of revenue, more or less important, in more than half of the states.

The United States income tax: General description. In the present federal income tax (in effect in 1948), the law distinguishes between individuals, corporations, and partnerships. All residents of the United States are in general subject to the tax upon all their taxable incomes. Nonresidents are generally taxable upon income derived from sources in the United States.

As regards the individual income tax, the law starts with the *gross income,* from which certain specified *deductions* are allowed, leaving the *net income,* which is the base of the tax. From the net income a personal *exemption* may be subtracted, amounting to $600 for each taxpayer and $600 for each person dependent for support upon the taxpayer. There are also $600 exemptions for taxpayers who are 65 years of age or are blind.

On the base thus determined, two taxes are then calculated: (1) a *normal tax,* at the rate of 3 per cent, and (2) a *surtax,* calculated by applying a progressive scale of rates from 17 per cent on the first $2,000 of the net income less exemptions to 88 per cent on the amount of this base in excess of $200,000. From the sum of normal tax and surtax certain percentage deductions are then allowed.[1] Certain credits may be deducted in special cases.

A novel feature of the 1948 revenue act was permission to married couples to combine their respective incomes, deductions, and exemptions, calculate the tax on half this base, and multiply this tax by two. This generally brings one of them into lower surtax brackets.

The law requires withholding by employers of part of the tax on wages and salaries.

By using as the base the taxpayer's advance estimate of his current year's income, the law is on as pay-as-you-go basis.

Corporations are separately taxed, by means of a normal tax and a surtax. The former is at rates from 15 per cent to 24 per cent on corporations having taxable net income of $50,000 or less; above $50,000, the rate is 24 per cent. The surtax is on a base reduced by

[1] In the 1948 revenue act these deductions are: on the first $400 of tax, 17 per cent; on the amount over $400 but not over $100,000, 12 per cent; and on the amount over $100,000, 9.75 per cent.

certain credits and is at rates from 6 per cent to 12 per cent on corporations having surtax net income up to $50,000; above $50,000, the surtax rate is 14 per cent. The law contains many complex, and frequently arbitrary, provisions affecting the definition of taxable income.

Partnerships as such are not taxable on their income. Their net income, whether actually distributed or not, must be allocated to the several partners and included by each in his individual income.

Individual income tax base. No income tax ever succeeds in reaching all the people. Personal exemptions, credits for dependents, and similar provisions result in freeing a large part of the population from income tax obligation. Thus in the United States returns were filed on the incomes of 1942 (the latest year for which the final official figures are available) by some 37½ million individuals, of whom about 28½ million had taxable incomes. With a population of 134⅔ millions in 1942, this means that returns were filed by less than 28 per cent of the population and income tax paid by about 21 per cent. Making allowance for married couples and minors and other dependents, these 28 million taxable returns may represent something like 78 million persons. This would mean that about 58 per cent of the population bore the burden of the federal income tax. The revenue act of 1948, by raising the personal exemptions to $600, removed some 7,400,000 taxpayers from the rolls. This reduced the above percentage from 58 to about 45.

Corporation dividends. The taxation of the entire net income of corporations may be regarded as double taxation to the extent of the part that is distributed in dividends to stockholders, since these dividends are then taxable as individual income. Until 1936 this was recognized, and a partial correction was made by requiring individuals to pay only the surtax (not the normal tax) on their dividend income. The effect of the present law is to discriminate against income from investment in the stock of corporations.

A simple correction of this double taxation would be to allow corporations to deduct dividends in calculating their taxable net incomes, just as they now deduct interest. There are however certain practical difficulties. Any other perfect correction would be very complicated, owing to the complicated rate structure of both individual and corporation taxes and the personal exemptions and credits allowed to individuals. A plan that combines simplicity and a reasonable approach to equity is to tax the stockholder on all his

income, including dividends, but allow him a credit determined by multiplying his dividends by the rate of the individual normal tax. There would remain double taxation only as to the surtax.[1]

Income tax rates. Beginning with the First World War there has been a revolutionary increase in income tax rates. In Great Britain, for example, during the fifteen years before the First World War, the basic income tax rates varied between 3½ and 6¼ per cent, with a super-tax of 2½ per cent after 1910. The present United States income tax was inaugurated in 1913 with progressive rates from 1 to 7 per cent. Today income tax rates have everywhere soared to hitherto unimagined heights. In the United States the present maximum rate (normal tax and surtax) of 91 per cent, though somewhat reduced in practical effect by the revenue act of 1948, is the highest tax rate known to modern history, and to this must be added additional income taxes in more than half of the states.

The practicability of such rates has come under serious question, especially in the United States. Even with the best possible administration, which the United States confessedly does not have, the success of an income tax must depend materially upon a prevailing spirit of fairness and coöperation on the part of the taxpayers. So long as this spirit is general, the tax administrators and the courts can cope with the actual criminal evaders. When this spirit disappears, and the majority become lawbreakers, then no administrative machinery that human intelligence can devise will be able to master the problem. This condition is threatened if rates are made too high.

It is possible also for income tax rates to be so high as to discourage investment in industry. This matter is discussed at greater length in the next chapter. For the present it is sufficient to point out that, with less investment and less industry, there will be smaller taxable incomes. The base of the income tax is thus reduced, and therefore the revenue. There is reason to believe that in recent years in the United States the federal income tax revenue has thus been less than it would have been with lower rates. In other words, the rates have been pushed beyond "what the traffic will bear."

Elasticity of the income tax. In the early days of the income tax,

[1] See The Committee on Postwar Tax Policy, *A Tax Program for a Solvent America*. New York: Ronald Press Company, 1945.

it was rightly regarded as a peculiarly elastic source of revenue, adaptable to normal variations in fiscal needs, and a tower of strength in time of war or other national emergency. The recent trend toward very high rates has seriously impaired the elasticity of the income tax.

Moreover this tax is peculiarly sensitive to general economic conditions. This was rudely disclosed with the coming of the depression after 1929. The financial exigencies of government led to increased demands upon the income tax, to which the tax, with its existing high rates, was not able to respond. In the United States, for example, even with increased rates, the yield of the income tax fell off disastrously. In 1930 the income tax produced 2,410 million dollars, contributing 58 per cent of the total revenues of the federal government. In 1933 its yield was only 746 millions, being 36 per cent of a total revenue which had itself dropped to less than half the level of 1930.

Consumption taxes. Next in importance to the property tax and the income tax comes a group of taxes whose central characteristic is the exaction of a tax contribution from the consumers of certain goods in proportion to their consumption. These taxes appear in a variety of forms. For the purpose of our study, we may divide consumption taxes into two main groups: (1) excise taxes (including sales taxes and business license taxes) and (2) customs tariffs.

Excise taxes. An excise is in general a tax levied upon the manufacture or sale or consumption of commodities, or the exercise of an occupation, or enjoyment of a privilege, within the limits of the jurisdiction imposing the tax. What will chiefly interest us are the taxes levied upon the domestic manufacture and sale of commodities, such as are commonly known in the United States as "internal revenue taxes" when imposed by the federal government and "sales taxes" when imposed by the states.

Excise taxes on manufacture are generally used by national rather than by local governments. This is principally because of the tendency of such taxes to drive industry outside the borders of the taxing jurisdiction. In the United States, the several states have generally been deterred from extensive use of excise taxes by these considerations, coupled with the constitutional prohibition of customs tariffs and of excises burdening interstate commerce. The states are not prevented, however, from levying moderate excise taxes upon industries catering to a strictly local market or upon local or

retail sales, though caution has always to be exercised lest the people be driven to make their purchases outside the state.

One of the most striking features of public finance following the First World War was the rapid and widespread development all over the world of sales taxation, forced by the fiscal needs of governments harassed by increased demands for expenditure and dwindling sources of revenue. Among the American states there has been a really extraordinary development of this form of taxation. The excise tax on gasoline is now universal. Excise taxes or license taxes are imposed upon sales of liquor by all states which do not prohibit its private sale. As to consumption generally, by the middle of 1948 more or less general sales taxes were in force in some twenty-seven states.

The customs tariff: General. Customs or tariff duties are taxes imposed upon those who bring goods across the frontier, the base being the goods so moved. The Constitution of the United States prohibits all duties on exports and also, in practical effect, forbids the levy of any customs duties by the states. From the beginning of the American nation in 1789 to the Civil War, the tariff was almost the only tax imposed by the national government, and from then to 1913 it divided honors about equally with the excise or internal revenue taxes. Today the tariff yields about one-tenth of the federal revenue.

Economic effects of the tariff. It may help the reader to grasp the salient characteristics of the tariff, together with their immediate consequences, if we set up a simple numerical example. Let us suppose that a certain kind of cloth is being imported, free of duty, and sold in the home market at $1.00 a yard and that 1,000,000 yards are sold in a given year. We assume that the cost of making this cloth at home is $1.50 a yard. Obviously the reason for the import is that this is the cheapest way to obtain the cloth, and of course there will be no home manufacture. The simple situation resulting is shown in column 1 of the table on the next page.

Now suppose that next year a tariff is in effect at the rate of 10 per cent, which will make the cloth cost $1.10 in the home market. As this price is still less than the home cost, there will be no home manufacture, and all the cloth consumed will still be imported. In accordance with the normal demand curve, other circumstances being unchanged, we assume that the amount consumed will fall off somewhat with the increase of price and that 900,000 yards will be

taken in this year. The result is shown in column 2 of the table below. Consumers pay $990,000 for cloth which, were it not for the tariff, they could have obtained for $900,000, and this additional cost goes to the government in revenue. Here is obviously the case of a simple tax on imports.

	1	*2*	*3*	*4*
	No tariff	10 per cent tariff	50 per cent tariff	75 per cent tariff
Number of yards imported	1,000,000	900,000	300,000	——
Number of yards domestic	——	——	300,000	600,000
Total number of yards consumed	1,000,000	900,000	600,000	600,000
Price per yard to consumers	$1.00	$1.10	$1.50	$1.50
Cost to consumers	$1,000,000	$990,000	$900,000	$900,000
Additional cost because of tariff	——	90,000	300,000	300,000
Revenue to government	——	90,000	150,000	——
Additional cost in excess of revenue	——	——	150,000	300,000

Next year a 50 per cent tariff is in effect, and the imported cloth will be sold, if at all, at a price of $1.50. This price is just equal to the cost of making the cloth at home, and the foreign and domestic products are in competition on an equal basis. Presumably they will divide the market between them in some ratio, which for the sake of simplicity we may assume will be an equal division. The result is shown in column 3 of the table, assuming a still further drop in the amount taken because of the rise in price to $1.50. Consumers now take 600,000 yards, paying for them $300,000 more than they would have paid for this quantity had there been no tariff. This additional cost, however, does not now all go to the government; the government collects 50 per cent upon the 300,000 yards imported, which gives it a revenue of $150,000, but it gets no revenue from the 300,000 yards of domestic manufacture. We meet here for the first time the phenomenon of *protection,* by which we mean *encouragement given to a home industry by means of a tariff.*

Now let us proceed to the last case, that of the *prohibitive* tariff. A duty of 75 per cent being in effect, the foreign cloth can be sold only at a price of $1.75. Since the home industry can furnish the cloth (at least 600,000 yards according to our assumption) for $1.50, there will be no possibility of selling the imported article. This situation is shown in column 4 of the table. So far as the consumers are concerned, the situation is the same as in the previous case, except

that all their purchases are of the domestic cloth. The government, however, now gets no share at all of the $300,000 additional cost imposed by the tariff upon the consumers. This is the extreme of protection.

To the reader who has fully grasped the meaning of the foregoing example, the most important principles of the customs tariff will now appear almost as a matter of course. They may be summarized in the following outline (the reader will find it helpful to refer again to the numerical example above):

(1) A tariff will cause the price of the dutiable article to rise in the home market by about the full amount of the duty, up to a duty which is equal to the difference between the home and the foreign unit cost of production.[1] Above that the duty is prohibitive, and the further result depends on whether there is competition or monopoly in the home industry. If competition prevails, further increments of duty have no effect on the price; under monopoly conditions the price may rise by more than the difference in unit production cost, up to the full amount of the duty.

(2) A tariff imposed at a rate less than the difference between the home and the foreign unit cost of production will have no immediate effect on home industry and will produce a revenue equal (disregarding costs of administration, etc.) to the added burden placed upon consumers.

(3) A tariff equal to, or greater than, the difference in unit production costs gives protection, and the additional burden placed upon consumers will be in excess of the revenue obtained by the government. If the tariff is in excess of the difference in unit production costs, it is prohibitive, in which case the additional burden on consumers is accompanied by no revenue whatever to the government. Stated in slightly different words, any tariff which gives protection to domestic industry necessarily causes a loss to consumers greater than the gain obtained by the government.

(4) The additional burden placed upon consumers by a tariff which gives protection will be equal to the difference between the home and the foreign cost of producing all the goods (domestic and foreign) consumed, except when a prohibitive tariff accompanies a

[1] The word "about" is introduced in this sentence in order to allow for the possibility of a change in the foreign cost of production resulting from the decreased purchase of the imported goods. The practical importance of this would generally be slight.

monopolistic condition in the home industry. In that case, the consumers may be burdened by an amount in excess of the difference in the entire cost of production.

The revenue tariff. The traditional country of free trade was Great Britain. From the reform of her tariff system in the middle of the nineteenth century until after the First World War, Great Britain followed a policy of tariff for revenue with virtually no protection. In order to make her tariff truly a revenue producer without disturbance from protective influence, Great Britain adopted certain devices which are universally effective for this purpose.

In the first place, her duties were placed upon a comparatively small number of articles, these being generally commodities which, in the nature of the case, could not profitably be produced in Great Britain, even in the face of high tariff duties. Such are tea, coffee, sugar, cocoa, spices, various fruits, and other tropical products.

The second device for safeguarding the revenue tariff from defeat through the checking of imports and the development of home industries is to impose an internal excise duty on each homemade good to match the tariff duty on the imported article. Finally, there is the extreme safeguard of prohibiting outright the home industry, which was employed by Great Britain in the case of tobacco growing.

By such means as these it is possible to build up a tariff system of extreme simplicity, easy to administer, having no disturbing influence upon the country's trade and industry, taking nothing from the people beyond what goes into the public treasury, and withal yielding a handsome revenue.

Fiscal weakness of the protective tariff. Few countries outside Great Britain, however, have chosen this kind of tariff. Virtually everywhere protection is consciously sought to a greater or less degree, and the employment of devices to neutralize the protective tendency of duties imposed for the sake of revenue is rare. At the extreme protective pole stands the United States, where, in spite of the great importance of the tariff as a source of revenue, the protective purpose has always been strong or even predominant.

Throughout the history of the United States, the tariff revenue has been notoriously irregular, out of relation to governmental needs, and not amenable to legislative control. Since the tariff is looked upon as an instrument for giving protection to certain industries, its rates cannot be raised or lowered to meet fiscal con-

ditions without opening up the whole political conflict over protection. Furthermore the relation between tariff rates and tariff revenue is extremely complex, and it is impossible to predict accurately the effect of a change. By checking the flow of imports, an increase of rates may diminish revenue instead of increasing it; a reduction of rates may increase revenue.

Death taxes: General description. The death of property owners gives to government, from time to time, a convenient opportunity to impose taxes upon the transfer of property of every description, an opportunity quite generally utilized by sovereign governments the world over. Death taxes are at present employed in the United States by the federal government and by all but one of the states (Nevada).

The typical modern *inheritance tax* contains the following technical features. At the outset property left to certain religious, educational, or charitable institutions is generally entirely exempt from the tax. Of the remaining taxable estate, a certain amount is usually exempt, this amount varying according to the degree of relationship between the decedent and the beneficiaries. Above the exempt minimum, the tax is calculated at rates which are usually doubly progressive: (1) the rates start low for bequests to the closest relations and increase step by step as the degree of relationship becomes more remote, and (2) within each class of beneficiaries, the rate is progressive according to the amount transferred. Among the states of the United States the top rate on bequests to direct heirs varies all the way from 1 per cent to 44 per cent; on bequests to the most remote relations and strangers, the two highest state rates are 60 per cent and 80 per cent respectively.

A tax levied upon the entire estate, without relation to the amounts of the distributive shares or the relationship of the beneficiaries, is called an *estate tax*. As an example we may take the United States estate tax, as in effect in 1948. Without attempting to summarize the extremely complicated provisions of the statute, we note that, as regards resident decedents, there is an exemption of $60,000, with progressive rates running from 3 to 77 per cent.

Death tax rates. It will be observed that the rates of the American death taxes—especially, the federal estate tax—have recently been pushed to very high extremes. This increase raises problems similar to those we have noted in connection with the income tax.

There is strong temptation to take advantage of all available

business arrangements and legal devices whereby death taxes may be escaped while still keeping within the letter of the law. There is, for example the possibility of limited escape through gifts before death, a door which no development of gift taxes is likely completely to close. The higher the rates, the greater is the inducement to resort to such avenues of escape.

There is also the effect of high tax rates upon saving and investment. This matter will be discussed at more length in Chapter xxv. Suffice it here to point out that the rates of death taxation may be so high as both to discourage saving and to prevent the continued existence of the large and medium estates which are the base of the tax. This tax, like the income tax, may be imposed at rates "higher than the traffic will bear."

Business taxes. It is scarcely possible to formulate a clean-cut definition of *business taxes* as a separate form of taxation. All taxes paid by business are not "business taxes." The owners and operators of businesses are in fact subject to most of the important forms of taxation which we have studied. Business property is subject to the property tax; business income, to the income tax.

There is however a feeling that engaging in business is to some extent a privilege, that business is conducted under the protection of the state, and that a special tax obligation is thereby created, and there is frequently a certain residue of taxation paid by business concerns that may be regarded as a business tax.

The clearest, though by no means the most important, examples are the special license taxes imposed as a condition of engaging in certain occupations, such as retail liquor selling, taxicab operation, insurance, etc. In many such cases, the regulatory purpose is more important than the search for revenue. Certain American states, especially in the South, have developed rather extensive systems of business license taxes, productive of substantial revenue, though now tending to decline in importance.

The most important business taxes are those imposed by the several states upon corporations. It is illogical to restrict business taxation to incorporated concerns. The reason for so doing is doubtless that most substantial business concerns are incorporated, and that incorporation makes them easier to reach. Some states do have, in addition, their special taxes on unincorporated businesses.

The most common taxes on business corporations are based on their capital stock, or their gross receipts, or their net income.

When the location of the corporation's property or the conduct of its business extends over more than one state, it is necessary to determine that part of the tax base which is properly ascribable to the taxing state. For this purpose, the laws generally contain rules of allocation, based usually upon location of physical property, payment of salaries and wages, or place where sales are made, or a combination of these bases.

Very often these taxes are in lieu of other state or local taxes (especially property taxes) to which the business corporation would ordinarily be subject. This is common in the case of the public utility corporations, particularly those that do an interstate business. Here is one of the instances where the tax cannot be regarded as wholly a business tax.

The subject of business taxation is a broad and complex one, for the details of which the reader is referred to the special treatises on taxation and public finance.

The poll tax. This is a tax, generally at a uniform rate of one to three dollars, upon every adult person, or every voter, or some other broadly inclusive class of the people. It is expensive to collect, except from those who are already subject to the property or income tax, and while it is (on paper at least) a recognized part of the American local tax system, it is generally honored more in the breach than in the observance, and the revenue derived from it is usually small.

Shifting and incidence: General principles. In drawing conclusions about taxation and its economic effects, we have always to remember that the burden of a tax does not necessarily rest upon the one who makes payment to the taxing government—the immediate payer, as we call him. It is necessary to know whether any particular tax is or is not shifted, and to understand the processes by which tax burdens may be passed on, in whole or in part, to others. Obviously the person who finally bears the burden is more to be considered than he who pays only to shift the burden to others.

A tax is or is not shifted according as it gives or does not give the immediate payer a power to raise the price of some commodity or service which he sells. It follows that the analysis of shifting is a problem of applying the general principles of price determination which we have previously studied. Now prices are determined under a variety of conditions: higgling and bargaining, auctions, pure competition, monopoly, and imperfect competition. In each of these market situations, the introduction of a tax on sellers or buy-

ers has its peculiar and distinctive effects upon their action. In general, a tax which increases marginal costs tends to be shifted, while a tax falling on some form of producers' surplus tends not to be shifted.

Any extended investigation of the specific problems of tax shifting cannot here be undertaken, but the general principles may be illustrated by a few simple examples.[1]

Incidence of excise taxes. The excises illustrate the operation of shifting in its simplest form. The United States internal revenue tax upon cigars is imposed upon all manufacturers of cigars at a certain rate per hundred and is thus an addition to the normal unit cost of production. Manufacturers are therefore able to increase the selling price by practically the amount of the tax and so to pass the burden on to the wholesale dealers to whom they sell their product. These in turn transmit the tax to the retailers, who pass it on in the prices charged the consumers.

Sometimes an excise tax may be imposed upon the dealers, wholesale or retail, instead of upon the manufacturers, or upon the consumers themselves at the time of purchase of the taxed articles, the law requiring the dealer to collect the tax and pass it back to the government. Whoever may be the immediate payer, the price paid by the consumer is of course increased by the amount of the tax.

When the excise tax is based upon an industrial occupation, instead of upon a commodity, the same result follows. The tax becomes an added cost of conducting the particular occupation and is passed on to the consumers by means of a higher price for goods or services. The retail liquor seller's license is a familar example.

The customs tariff. The shifting and incidence of the customs tariff have been illustrated in considerable detail earlier in this chapter. They are governed by the same principles that apply to the excise taxes.

Taxes on economic rent and land value. The simplest illustration of a tax that is not shifted is the tax upon the landowner's income from economic rent. Having in mind the nature of economic rent, we readily reach the conclusion that a tax upon economic rent gives the landowner no power to increase the rent he may be charging. The tenant (assuming he is already paying the full economic rent) can pay no more rent without being forced to

[1] For a fuller treatment of this subject, see Fairchild, Furniss, and Buck, *Elementary Economics*. New York: Macmillan Company, Fifth edition, 1948, Chapter XLV.

operate at a loss. Rather than do this he will leave this land and either go out of business or betake himself to the marginal land, where there is no rent and therefore no tax on rent.

On the other hand, neither the tenant nor the landowner who is using his own land is given any power to charge a higher price for his crop by a tax on economic rent. The price is determined by demand and the cost of production on the no-rent land, neither of which is affected by the tax in question. We conclude that a tax on the economic rent of land is not shifted but is borne by the immediate payer, the landowner.

One necessarily reaches the same conclusion as to a property tax upon the value of land, since the value of the land is the capitalized value of its economic rent. A property tax on land can certainly give the owner no power to sell the land at a higher price; nor does it give him power either to exact more rent from a tenant or charge higher prices for agricultural products, since the marginal land, having no economic rent and so no value, is no more affected by a property tax than by a tax on rent.

Capitalization of taxes. The tax on economic rent or on land value furnishes the simplest illustration of the principle of "tax capitalization." If an acre of land is yielding an economic rent tax-free of $10 a year and the rate of interest at which such incomes are discounted is 5 per cent, the land will be worth $200. If now, assuming no other material change in the tax system, an annual tax is imposed on economic rent at the rate of 10 per cent, the landowner will have left an income of only $9.00 a year, which, capitalized at 5 per cent, makes his acre of land worth $180. Taking 10 per cent of the annual income is equivalent to taking in one lump sum 10 per cent of the value of the land. No one will buy this land for more than $180, and the purchaser, receiving $10 in rent and paying $1.00 in tax, obtains 5 per cent on the investment, the same as he would have received if he had paid $200 for an acre not subject to the tax on its rent.

Of course precisely the same result follows the imposition of an annual property tax on the value of the land. Taxes on economic rent or land value thus tend to rest as a permanent burden upon those who owned the land at the time of the imposition of the taxes and not to be passed on to later purchasers. The principle of capitalization applies not only to land but to any other capital instrument where it is not possible to shift the tax—for example, a tax on bonds.

CHAPTER XXV

Principles of Taxation

An attribute of sovereignty. Without revenue, no government is able to perform its allotted functions. And if its revenue comes, not from its own inherent power to tax, but as a grant from some other authority, the government is not sovereign. The power to tax is thus one of the attributes of sovereignty. These important principles are well illustrated by the history of the American people.

During the War of the Revolution, the government of the whole people, such as it was, was in the hands of the Continental Congress. The American colonists were always sensitive to taxation. It was a sore point between them and Great Britain, and when the Revolution released them from British taxation, they were loath to give to their own national government adequate taxing powers. The Continental Congress had thus no power to tax the people. Its revenues consisted of grants from the several colonies or states, or taxes imposed only with their consent and for whose enforcement the Continental Congress was without authority. The fiscal affairs of the Revolution and indeed the whole direction of the war by the Continental Congress presented, in consequence, a record of impotence and ineffectuality. Actually the war was conducted without the services of a sovereign national government.

After the war, the states were still reluctant to surrender any of their sovereign control of taxation to a central government. They set up, under the Articles of Confederation, a government whose brief inglorious record of impotence was chiefly due to lack of the essential power to impose taxes on its own authority.

Gradually the people came to understand that they could never go forward as a great and prosperous nation, taking advantage of their fabulous heritage of natural resources, without a real sovereign national government whose sovereignty included the power to tax.

Sovereignty and taxing power in the United States. So in 1789 the American nation was born under the Constitution of the United

States. For the first time there was a national government with undisputed power to tax the wealth and the incomes of the people without the consent of any state.

This is not to say that the new national government's sovereignty or its taxing power was unlimited. What the people did was to set up a federal form of government. Certain limited taxing powers were conferred upon the national government and certain corresponding limitations placed upon the taxing authority of the states. The states remained the residual possessors of all sovereign powers of taxation not granted, specifically or by implication, to the national government.

As to taxing powers, the principal limitations of the Constitution are these: (1) neither the national government nor the states may impose taxes upon exports; (2) the national government has the exclusive power to tax imports (customs tariff) and to impose taxes that burden interstate commerce; (3) direct taxes by the federal government, except income taxes (since 1913), must be so levied that the amounts collected in the several states will be in proportion to the populations of the states; (4) indirect taxes must be uniform throughout the nation (meaning only geographical uniformity); (5) the states may not tax exports or imports, nor lay taxes burdening interstate commerce, and they are further limited by the fact that treaties of the national government take precedence over state laws.

Judicial interpretation later developed another very important constitutional limitation, which is sometimes called the "immunity principle." This means that the federal government is generally forbidden to tax an instrumentality of a state government, and likewise that as a general rule no state may tax an instrumentality of the federal government. The first announcement of this principle was in the decision of Justice Marshall in the famous case of *McCulloch* v. *Maryland* (decided in 1819), in which the state of Maryland was forbidden to levy a tax upon the notes of the Bank of the United States, a federal institution. Similarly the United States Government is not permitted to tax the interest on bonds and notes of the state and municipal governments.

In a federal form of government, this mutual immunity is essential to the preservation of the sovereignty of both the national government and the states. There is no such principle as between the states and the local governments, since the latter are not sovereign.

The purpose of taxation. The purpose of government is to serve the welfare of the state or of the individuals who compose the state. These two phrases are not identical; indeed they represent two distinct and conflicting schools of thought. The first—which has been called the German school and which is represented by all totalitarian states, whether monarchic, communist, socialist, or fascist—holds that the state is above all. Its power, prestige, and continued advancement are the prime objects of all policy; the welfare of the individual members, if considered at all, is subordinated to the interests of the state.

The second school of thought used to be called the English school in contrast with the German. Its attitude is in general that of democracy. According to this school, the interests of the state are secondary to those of the individual. The welfare of the individual is the ultimate goal; the state exists only to serve the interests of the individuals, and the organization and operation of government are to be directed accordingly.

A free economy, as we have envisaged it, can exist only under this second concept of government. Since, as has been stated, this book deals with the economy of freedom, it is not necessary for us to enter into discussion as to the relative merits of these two concepts of the state and the aims of government. We shall proceed from the premise that in the American democracy the state—the government—is presumed to exist and function, not for its own sake, but to serve the interests of the individuals who compose it.

From this beginning we proceed naturally to the notion that the purpose of taxation, in the broadest sense, is to promote the welfare of the people. More specifically, we may say that the primary purpose of taxation is to raise the money necessary to pay the costs incurred by the government in the performance of its functions; *i.e.*, in rendering its services to the people.

In contrast with this purpose, taxes may be imposed, not for the sake of raising money, but as the means of accomplishing some other objective. Such objectives are called the nonfiscal purposes of taxation, and we shall give them some attention later. For the present, we shall be thinking chiefly of taxation as imposed for the primary purpose of paying the cost of government.

The *quid pro quo* of taxation. In general. Taxes are always a burden upon their subjects and so, all together, upon the whole community. Occasional efforts to picture taxation as in some way

a blessing have never succeeded. What is a blessing is the flow of governmental services for which taxation pays the cost. Taxes are the price the people pay for the services of government. Of course it cannot be asserted that every activity of government does actually confer a benefit upon the people. It is possible that, through ignorance or special influence, an act of government may confer no benefit, or may even be harmful.

But the general presumption is that government services are beneficial, or at least are what the people want, and that they are therefore the *quid pro quo* for taxation. Taxes are burdensome; government services are beneficial. How do we strike the balance?

If we may be permitted for the moment to employ, in a figurative sense, the marginal concept which we have found useful in other connections, we can say that the optimum amount of taxation, from the viewpoint of the welfare of the people, is reached when the marginal benefit of government services rendered just equals the marginal sacrifice of taxes collected. At this point the people would be equating the marginal utilities of government services with the other things for which they might spend their money. If government did less than this, the people could gain by paying more taxes in order to get more government service and correspondingly less of other enjoyments. If government did more than this, the people would not be getting their money's worth in government service for their marginal tax payments. They would receive more satisfaction if their marginal dollars were spent for other things.

While this sort of approach may serve the purpose of stating the problem, it will not get us much farther. Actually the people as a whole do not experience pain from paying taxes, or satisfaction from government services. It is only individual persons who experience pain and pleasure. The burdens and satisfactions of the whole community can mean nothing except the sum of the corresponding reactions of individual persons. The problem of the *quid pro quo* of taxation thus relates to the benefits of government services and the sacrifices of tax payments as experienced by individuals. For the further analysis of this problem, we have to inquire into how both government services and tax levies are distributed among the people.

Distribution of the burden of taxation: How "ought" taxes to be distributed? For generations students of taxation have been wrestling with the question of how the burden of taxation "ought" to

be distributed among the people. This is clearly an ethical question. Science does not undertake to answer ethical questions, and the science of economics cannot tell us how taxes "ought" to be distributed. What we can do is to make acquaintance with some of the principal theories or formulas that have been proposed for the distribution of taxation and seek to discover how they would work and what their consequences might be.

Obviously one's notion of how taxes "ought" to be distributed will depend on his conception of the basic purpose of government. We shall limit our present discussion by adhering to our premise that the government exists to serve the individual members of the community, rather than for its own sake. The goal of the ideal formula for tax distribution will then be the welfare of the individual members of the state.

Tax distribution formulas: Benefit or cost. Some persons have believed that the costs of government should and could be distributed on the basis of the benefits received by the individual members of the group, or on the somewhat similar basis of the cost to the government of the service rendered to each person. These are in general the rules that determine prices charged in business dealings, and we have seen that they apply also to industrial enterprises conducted by government.

But when we come to the general services of government, performed for the common good rather than for the particular benefit of individuals, there is generally no way of measuring either the benefit or the cost to each person. And even if some measurable relationships were ascertainable, the government could collect only from those who were able to pay, though its general (primary) services must be rendered for all without discrimination. Evidently neither benefit nor cost is capable of providing a workable general rule for the apportionment of taxes.

Ability to pay: What does it mean? Probably the tax distribution formula in greatest favor today is taxation according to "ability to pay." This has a good sound. What could be fairer than to collect from each according to his ability? To many the meaning of this formula seems so clear as scarcely to require definition, and that taxes "ought" to be apportioned according to ability seems axiomatic. Once we start inquiring, however, into the precise meaning of taxation according to ability to pay, we shall find the notion by no means so simple and axiomatic as is sometimes assumed.

The first step is fairly easy. Although some scholars have sought to distinguish between the sacrifice of paying taxes and the faculty or capacity to pay taxes, the ability theory is almost always expressed in terms of the sacrifice suffered by the taxpayer. Taxation according to ability to pay may be taken to mean (negatively stated) taxation according to sacrifice.

But when we proceed to ask what is the meaning of taxation according to ability or sacrifice, we discover that there is no generally accepted answer. The fact is that the writers are divided into some three or four schools of thought according to their several ideas of what is the object to be accomplished, and their definitions of taxation according to ability to pay or sacrifice differ accordingly.

Equal sacrifice: Basic assumptions. There is first the equal sacrifice theory. According to this formula, taxation according to ability or sacrifice requires that each taxpayer lose the same (absolute) amount of utility as every other taxpayer. Investigation of the kind of tax system that would produce this result is based upon the principle of diminishing marginal utility, with which the reader is familiar. It is assumed that everyone's estimate of money or purchasing power is subject to this principle; *i.e.*, the more money one has to spend, the less is the utility of one dollar to him. This may be expressed by the familiar curve of marginal utility, and it is further assumed that this curve has the same shape for all persons. In other words, the curve (such as *mn* in Fig. 30 on page 488) which shows for an individual, A, the marginal utility of money to him in relation to the number of dollars he possesses, will show also this same relationship for all other persons. Thus, if A has *Oa* dollars, the marginal utility of money to him is *ad;* if B has *Ob* dollars, the marginal utility to him is *be;* and so on.

Accepting these assumptions, we may conceive of a universal curve of the marginal utility of money, which would show the marginal utility of any quantity of money to any person, according to where he stood on the scale of amount of money possessed. Waiving for the moment the question of its precise shape, such a curve may be illustrated by the curve *mn* in Figure 30.

The problem then is to discover what kind of tax system would produce equal sacrifice. In particular, we want to know whether such a system would be proportional, progressive, or regressive. The answers to these questions depend on what is the precise pattern of the relationship between quantity of money possessed and marginal

utility of money; *i.e.*, on the precise shape of the curve of marginal utility of money.

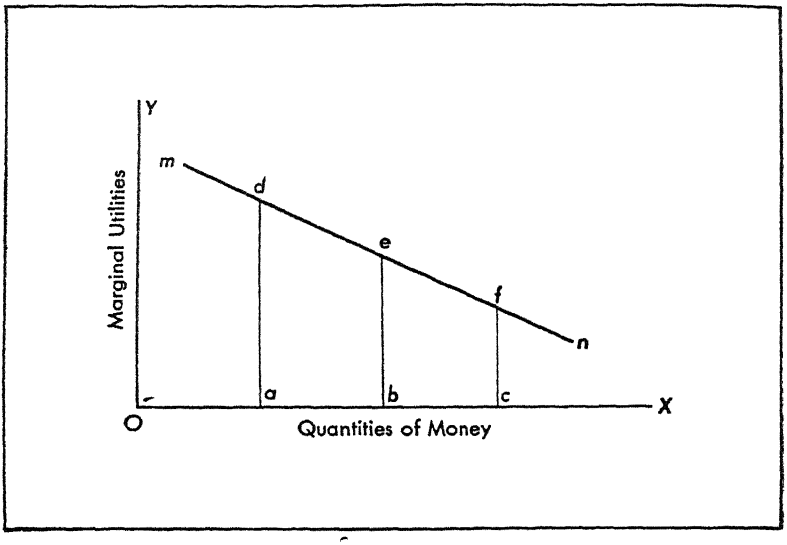


FIG. 30

As a first approximation, consider the curve *mn* in Figure 31. This curve has been, arbitrarily, so constructed that the ordinates measuring marginal utilities vary in inverse proportion with the ordinates measuring money possessed. The student who is mathematically inclined will recognize this as a rectangular hyperbola, whose equation is $xy = c$.

If now we assume, for the moment, that this is the curve that truly expresses the relation of the utility of money to the amount possessed, then it is clear that equal sacrifice will be accomplished by a proportional tax; *i.e.*, a tax at a flat (uniform) rate. If B has twice as much money as A, he pays twice as much tax; but the marginal utility of money to him is only half what it is to A. Hence the sacrifices of these two men—represented by the two shaded areas—are equal. This is easily demonstrated geometrically, the base for B being double that for A, while B's median altitude is half that of A's.

If, on the other hand, the curve that truly represents this relationship should actually be steeper than the rectangular hyperbola,

then B's sacrifice under a proportional tax would be less than A's, and the goal of equal sacrifice would require progressive taxation. If the curve should actually be less steep, regressive taxation would be required. To solve our problem we must know the shape of this curve.

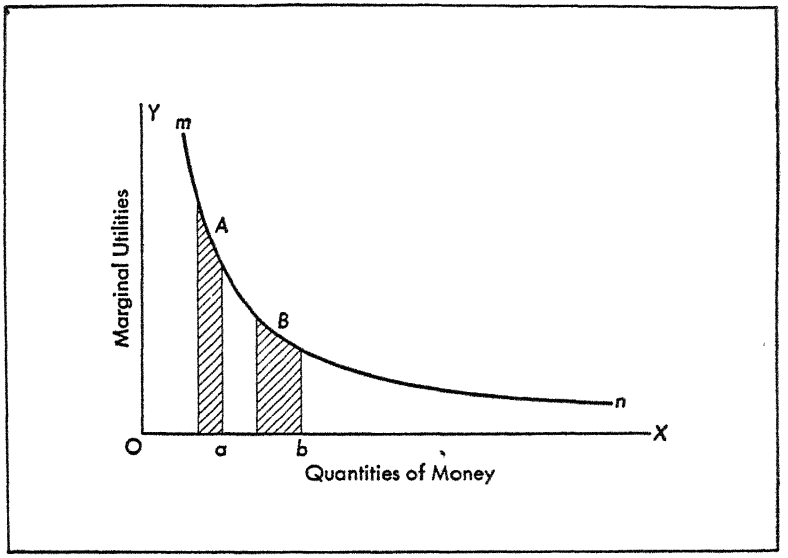


FIG. 31

Testing the assumptions. How far does all this get us toward our goal—the discovery of the kind of tax system that would produce equal sacrifice? Let us examine some of the assumptions which we have had to make.

(1) We assumed a universal curve of the marginal utility of money. Preliminary to this, we had to assume that there is a measure of the marginal utility of money. We have previously learned that it is possible to measure the marginal utility of a particular good to a particular person relative to, or in terms of, money. But it is very doubtful whether there is a corresponding measure of the marginal utility to this person of goods in general—which is what is meant by the marginal utility of money.

Moreover it is not enough to measure the marginal utility of money to a particular person. We must have a common unit that

will measure the marginal utilities of different persons. This appears to require an absolute measure of marginal utility. But we have learned that there is no such thing. Though A and B may have identical means, we cannot say that the marginal utility of money is the same to both. In fact, realizing the possibility—indeed, the certainty—of a host of differences in their respective personal situations, we can be sure that, except by rare coincidence, their marginal utilities would not be identical (even were there a way of measuring utilities of money). Still less can we assume that the marginal utilities of the different amounts of money possessed by A and B respectively would bear the same relation to each other as is borne by the utilities of these same amounts when plotted on the curve of one person.

(2) Not only had we to assume the existence of a general curve of marginal utility of money, but we further assumed knowledge of the shape of the curve. Without this knowledge our problem is indeterminate. Serious efforts have been made by a number of economists to determine the shape of the curve of marginal utility of money, but so far without success. It is doubtful if this quest will ever succeed. And considering the shaky assumptions on which the concept rests in the first place, it appears that, even if the answer sometime should be found, its practical usefulness would not be great.

This is not all. The discussion so far has been in terms of a curve of uniform slope throughout its entire length. This assumes that, whatever may be the relation between quantity of money and marginal utility at any point, it will be the same at every other point. But as we learned when studying the elasticity of demand and supply, this sort of relationship does not have to be the same for all quantities. The curve (assuming there is any such curve) might not be of uniform slope; some parts might be steeper than others. We must admit the possibility that, if we had the true curve, it might indicate progressive taxation for certain ranges of wealth or income, proportional taxation for others, and regressive taxation for still others. As a matter of fact, this may actually be the situation for all we know.

The other sacrifice theories. Thus far we have been dealing with the equal sacrifice concept of ability to pay. But this is not the only possible meaning. There are at least three other ideals, known respectively as proportional sacrifice, equi-marginal sacrifice, and

minimum sacrifice. All the sacrifice theories rest on the same basic assumptions which we have critically examined in connection with our study of the equal sacrifice formula. We therefore do not need to explore further these other theories.

Is sacrifice "gross" or "net"? One further observation of a general nature must be made. In our discussion thus far of the ability or sacrifice theory, we have considered the direct sacrifice of paying taxes, without giving thought to what the taxpayer gets in return in the form of benefit from government services. In other words we have considered the "gross" rather than the "net" sacrifice. Yet every taxpayer may be assumed to derive some amount of benefit from what the government does. By some taxpayer, it may be that the services are considered worth the tax; he would be willing to pay that much for them if they were not thus furnished by government. To everyone the net sacrifice of taxation must be at least somewhat less than the gross.

Of course the general services of government, for which taxes are collected, are not rendered specifically to particular persons—even less, sold to persons in return for definite tax payments—and it would be quite impossible to determine for each taxpayer the quantity of benefit from government services that should be deducted from the sacrifice of his tax payments in order to derive the net sacrifice he suffers. Yet it would seem that a theory of taxation according to ability or sacrifice would have to take account of this. What we have is a still further cloud of uncertainty over the concept of taxation according to ability to pay.

Conclusion. Summing up, it appears that the concept of taxation according to ability to pay rests upon basic assumptions which, if not contrary to fact, are at least so arbitrary and unrealistic as to make this concept both theoretically untenable and practically useless.

Many people—perhaps most people—probably have the feeling that the marginal utility of money decreases faster than the rate at which the amount of money increases; in other words, that there is a curve of marginal utility steeper than the rectangular hyperbola. As is said, a man with an income of $10,000 would feel the loss of $1,000 less than the man with an income of $1,000 would feel the loss of $100. Assuming equal sacrifice as the goal, this would point to progressive taxation. But this popular notion is extremely vague, especially as to its essential assumptions.

Just taxation. The rule of equity. To some it may appear that a sufficient rule for the apportionment of taxation would be to require that the apportionment be just or equitable. Probably everybody who conceives of government as in the interest of the individuals that compose the state would agree at once with the proposition that "taxes ought to be justly distributed." But this is of little help as a practical rule of tax apportionment. As soon as we attempt to formulate a plan of tax distribution according to the specifications of the rule of equity, we are confronted with the lack of any absolute standard of justice. What is equity? What is justice? There is no answer that will serve our practical problem. About all we can say is that, for any community, justice is that standard which has come to be generally accepted by the public opinion of the community. And even if we are ready to accept some particular standard, we find lack of any precise formula that would tell us how taxes should be apportioned. Attempts to apply the standard of justice usually fall back on the utility or sacrifice theories which we have just examined.

As regards the details of a tax or a tax system, it may sometimes be possible to criticize and reach helpful conclusions by reference to the test of justice in accordance with generally accepted standards. Thus, "double taxation" can generally be branded as "unjust." But we are on dangerous ground even here. Tax exemptions and subsidies, for example, which rationally would appear to offend the standard of justice, are frequently accepted by public opinion without serious misgiving on grounds of equity. And as a rule by which to establish a broad plan of tax apportionment, the rule of equity provides little help.

Progressive taxation: Introduction. In virtually all discussions of the problem of apportioning the tax burden, the question of progressive taxation looms large. Ability to pay and sacrifice theories are inextricably involved in assumptions or questions as to progressive taxation. And the same is true of the popular ideas of "just" taxation. The time has come when it will be worthwhile to give more searching attention to the nature and the consequences of progressive taxation. In this discussion we shall generally be using, not the strict technical definitions, but the more popular though less precise concept of progression in accordance with general economic power (see pp. 461–462, Chapter xxiv).

Is progression required by justice? We need not pause long over the claim that justice requires progressive taxation. We have observed the absence of any standard of justice that could make this claim mean anything practically. Most of the discussion of this matter only begs the question. Progressive taxation is advocated by many on the ground that it is required by the dictates of justice. But when we inquire what justice is in tax matters, we are likely to find that progressive taxation is assumed as part of the definition of justice.

Does ability to pay require progression? Is progressive taxation required by the rule of distribution according to ability to pay? We have already ventured the opinion that many people—perhaps most—would probably answer offhand, yes. But our analysis has shown both that the concept of ability to pay is a vague one and that it does not necessarily require progression. Some of the sacrifice theories require progression, others do or do not according to the assumptions that are made. We certainly cannot accept as an established general principle that taxation according to ability to pay or sacrifice requires progressive taxation.

The degree of progression. If there is no principle that progressive taxation is required, either by the rule of equity or by the formula of ability to pay, still less is there any principle that would determine the degree of progression that might be required by either rule. As to this—whether progression "should be" moderate or steep—we have no guide.

"Equalizing fortunes." What progressive taxation does do is to provide a force tending to modification of the existing distribution of wealth and income—"equalizing fortunes," as the saying is. That progression has this effect is inherent in its very definition. And obviously the higher the degree of progression, the closer will be the approach to equality.

Whatever one's opinion as to the desirability of this result, here, it would appear, is the only rational ground for advocating progression. Progressive taxation is "good" or "bad" depending on whether one desires or does not desire an approach to equality of fortunes brought about in this way. And the answer to the question of the degree of progression is to be found likewise.

Prevalence of progression. The notion of progressive taxation is a comparatively modern one, associated chiefly with income and death taxes. The income tax did not get under way generally in

Europe until around the middle of the nineteenth century, and in the United States, not until the second decade of the twentieth century. The significant development of the death taxes came, both in Europe and America, only in the twentieth century. Progression has been applied almost exclusively to these two forms of taxation.

Income taxes are employed by thirty-four of the states. They are all progressive, with rates running from 1 per cent to 15 per cent. The median of the highest bracket rates is 5 per cent.

All the American states but one—Nevada—have death taxes of some form. Many of these started with flat rates, only later introducing progression. Those that started with progressive rates have generally made the progression steeper with successive amendments. At the present time progressive rates are used in all but four of the states. The rates run from ⅘ per cent to 40 per cent, except for Minnesota, with 60 per cent, and Delaware, with 88 per cent. The median of the top rates is 13 per cent.

The use of progressive taxation by the states as a whole appears modest when attention is turned to the record of progressive taxation as employed by the federal government of the United States. The original income tax of 1913 imposed a normal tax, at the rate of 1 per cent on incomes of both corporations and individuals, and a surtax on individuals, with progressive rates from 1 to 6 per cent. From this modest beginning, progression advanced to its highest level in 1945, when the income tax rates were: corporations, 15 to 40 per cent (corporations had in addition several other taxes to pay); individuals, 23 to 94 per cent. In the fall of 1945, these rates were slightly reduced, making the top rate for corporations 38 per cent and for individuals 91 per cent (but with a 5 per cent deduction from the total computed tax).

The present federal estate tax had its origin at the time of the First World War, 1918. Developing from a comparatively modest beginning, its progressive rates now run from 3 per cent to 77 per cent.

Revenue aspects. From the revenue standpoint, progressive taxation is not so effective as is commonly supposed. Progression may collect very large amounts from wealthy individuals as compared with what is paid by those in the lower income groups. But there are so few of the former as compared with the latter that, in the aggregate, the bulk of the revenue must come from those in the lower and middle classes. For example, in 1942 the entire number

of those reporting net incomes of $25,000 or more was only 89,427, being only ¼ per cent of the 36,366,683 returns for that year. The aggregate of net incomes reported by this group made up less than 6 per cent of the total reported. Ninety-four percent of all individual net income reported in 1942 was received by persons having less than $25,000 each. Eighty-two per cent was in the hands of those having less than $5,000 each.

If a large sum is to be raised by income tax on individuals, the government must go where the money is. The bulk of the yield will have to come from the basic or "normal" rate which is paid by everybody. "If the normal tax is low, no degree of progression, however severe, can produce enough to provide an impressive total tax yield from individual incomes."[1]

This is clearly illustrated by the following table, which shows the amounts that would be raised respectively by the true normal tax (*i.e.*, the rate which applies to all incomes regardless of their size) and the true surtax (*i.e.*, the progressive rates) when applied to a total taxable income of 44.7 billion dollars at the tax rates that were in effect in certain years, as estimated by the Committee on Postwar Tax Policy.[2]

TAX YIELD FROM A TAXABLE NET INCOME OF 44.7 BILLION DOLLARS AT THE RATES APPLICABLE IN 1940, 1941, 1942, AND 1944, RESPECTIVELY

(In millions of dollars)

Tax	1940	1941	1942	1944
True normal tax	1,788	4,471	8,495	10,284
True surtax	1,925	2,704	2,799	2,743
Total	3,713	7,175	11,294	13,027

Although surtax rates were substantially increased from 1940 to 1944, what caused the very much higher revenue in 1944 was the increase in the (true) normal rate. This was 4 per cent in 1940, 10 per cent in 1941, 19 per cent in 1942, and 23 per cent in 1944. Of the increase from 3⅔ billion dollars in 1940 to 13 billions in 1944, 8½ billions came from increasing the normal rate and less than 1 billion from the progressive rates. And the total yield in 1944 came, four-fifths from normal tax, and only one-fifth from

[1] The Committee on Postwar Tax Policy, *A Tax Program for a Solvent America*. New York: Ronald Press Company, 1945, p. 37.

[2] *Ibid.*, p. 37.

progression, though the latter was at drastic rates, reaching 82 per cent in the top bracket.

As a matter of fact, this illustrates precisely what the United States had to do when, in 1941 and 1942, the income tax was called upon for large additional yields. Little was obtainable from the progressive rates. Instead, the surtax was made applicable to the first income bracket, having the same effects as raising the normal tax; thereafter this combined initial rate was further increased as stated above.

In the revenue act of 1942, the surtax rates ran from 13 per cent to 82 per cent, the highest being on the excess of income over 5 million dollars. If the progression had instead stopped with 52 per cent, which was reached on the taxable income bracket $22,000–$26,000, the loss of revenue would have been about 249 million dollars, or 2.8 per cent of the total individual income tax of that year. The entire system of progressive surtax rates from 52 per cent to 82 per cent could have been thrown overboard with a loss of only 2.8 per cent of the revenue.

A neutral tax system: Viewpoint. By those who recognize the dubious foundation of the various ability-to-pay and sacrifice theories and who, appreciating that the only rational purpose of progressive taxation is to reduce inequalities of wealth, do not favor the use of taxation for this purpose, there is offered another approach to the problem of apportioning the tax load. The goal is a tax system that is neutral as regards the distribution of wealth and income; one that leaves each taxpayer in the same economic position, relative to his fellows, that he occupied before taxation was imposed.

Narrowing the field of taxation. As a first step toward a neutral tax system, it is suggested that, among the government functions presently financed by taxation, there doubtless are many items whose benefit or cost could be allocated with reasonable approach to accuracy.

We have observed that it is quite logical for the government to charge prices for the goods and services offered to the people by government industrial enterprises; likewise, to collect fees where a charge is possible based on benefit received from government or cost occasioned to government by particular persons. The trouble with benefit or cost, as a general theory for distributing taxation, is that benefits or costs of the general services of government cannot be

apportioned to particular persons. Careful scrutiny of government services now supported by taxation would, however, disclose numerous items which could be financed by prices or fees on a cost or benefit basis.

Examples of a trend in this direction are the motor vehicle licenses and gasoline taxes, which are aiming toward, and in some cases rather closely approaching apportionment of the cost of the highways on the basis of personal benefit or cost. The fact that some of these charges are called taxes rather than fees is, as we have seen, immaterial. The same principle is illustrated in the increasingly common practice of financing the capital outlay, and sometimes also the operating costs, of newly constructed parkway roads, bridges, etc., by tolls collected from the users.

It is not meant to suggest that progress is all in this direction. There is, in fact, a trend also in the opposite direction. The field of government activity is being expanded by services rendered to particular persons and financed by taxation, even though these services could be provided by private enterprise, or, if furnished by government, could be financed by prices or fees collected from those who receive the benefit.

Thorough examination of all the services of government, followed by substitution of prices and fees for taxation when it was found practicable to allocate personal benefits or costs, would narrow the field of the general services of government and reduce by so much the magnitude of the problem of apportioning the burden of taxation.

However after all is said and done, it will always be true that the great bulk of the services of government are of a general nature, neither their costs nor their benefits being capable of allocation to particular persons. Here financing must be by taxation, and the problem of apportionment remains. This, it will be recalled, is inherent in the definition of a tax.

Proportional taxation. A tax system that is neutral as regards the distribution of wealth will be neither progressive nor regressive. The former tends to reduce relative inequalities; the latter to increase them. Neutral taxation will be proportional. Taking from all taxpayers the same fractional part of their resources leaves them in the same relative positions as before.

The neutral tax system does not necessarily exclude progressive taxes, or regressive taxes, or even equal taxes. It is the whole

system that is neutral, not necessarily the several separate taxes of which it is composed. The progressive nature of one tax may be offset by another tax that is regressive, so that the combined tendency of all the taxes in the system is neutral. Thus it is conceivable that the progressive income tax of an American state might be balanced by the regressive character of its sales tax. Even a tax distributed equally per capita, like the poll tax, might not cause the whole tax system to be regressive.

Herein lies the error of those who, starting with the premise that regressive taxation is "bad," condemn the sales tax because it is regressive. A particular tax cannot be so judged. It is the tax system —the complex of all the different taxes that are employed—that is to be judged progressive, or regressive, or proportional.

A principal difference between a progressive tax as part of a proportional system and a tax system progressive as a whole is that in the former case there is a limit to the degree of progression. When the progressive rates of certain taxes have become high enough or steep enough to counteract the regressive nature of other taxes, making the whole system proportional, then the limit of progression is reached. Under the theory of progressive taxation for the whole tax system, there is, as we have seen, no rational basis for any limit to the degree of progression.

Relation to ability to pay. It is not correct to say that a proportional tax or tax system has no relation to ability to pay, so long as one accepts any index whatever for measuring relative ability. Taking the very simplest premise—*i.e.*, that as a general rule ability to pay varies directly with the amount of wealth or income possessed—practically every modern tax (see Chapter XXIV) is positively related to ability to pay. The man with a large income pays more income tax than his friend of smaller income. The more property one has, the greater is his property tax. A sales tax is levied in proportion to purchases, which are on the whole an indication of relative incomes. The only tax that fails to take account of ability is the poll tax; it treats everybody alike, regardless of property or income or any other criterion of ability to pay.

The trouble with the conventional theories of taxation according to ability to pay is that they are not content with a mere direct relationship, but require that taxes should vary in exact proportion to ability to pay. It is here, as we have seen, that we get into difficulties, difficulties which the simple proportional tax theory escapes.

Summary and conclusion. The conclusion seems to be inescapable that there is no generally accepted formula for the distribution of the burden of taxation. The impracticability of apportionment according to benefit or cost is easily demonstrated and generally recognized by all those who have given thought to the matter. On the other hand, the principle of distribution according to ability to pay or sacrifice has attained wide popularity, being to many a sort of article of faith. Yet critical analysis shows this theory to rest upon concepts that are incapable of precise definition and assumptions that are arbitrary and unrealistic. Resort to the simple rule of equity gets us nowhere. And finally the rule of neutral or proportional taxation, though having much in its favor through simplicity and logic, probably does not enjoy wide popular acceptance.

The fact is that rules, formulas, principles, and theories of tax apportionment have generally not played a very important role in tax legislation. The formula has more often followed than preceded the construction of a tax system. Law-making bodies seldom start out with an accepted formula as to the proper distribution of the tax burden and then legislate to the pattern. Outside of academic dissertation, tax distribution formulas have played a role in general discussion chiefly as a means of rationalizing tax legislation already enacted or justifying some tax program that is being advocated.

In a self-governing democracy the adoption of a tax system, like all other forms of legislation, is in response to a complex of influences. Political considerations, the influence of minority pressure groups, and all the many factors that contribute to public opinion, play their part. Public opinion regarding taxation is without doubt influenced by the theories, vague and unscientific though they may be, which people entertain regarding the proper distribution of the tax burden. It is here that formulas of tax apportionment play their part and have importance, rather than as a direct guide to tax legislation.

Collateral effects of taxation. Of course what taxation accomplishes is the raising of money to pay the costs of government. But from the process, there come other consequences—sometimes foreseen or even intended, sometimes not—which are beyond this primary purpose and effect. Let us give attention first to those results that were not foreseen or intended, which we may call the collateral effects of taxation.

Geographical shifting of industry. Taxation may cause geographical shifting of manufacture and marketing. A tax by a state on manufacturing enterprises tends to force business to relocate in other states where the taxes are not so severe. It is for this reason that the American states have to be careful in their taxation of business. In choosing between places in the same state, a businessman contemplating the location of a factory will give attention to the relative severity of the local tax rates.

A state sales tax encourages the people to buy from dealers in other states and may force stores to move across the border into an adjacent state. This tendency is of course restricted by the necessity for retail stores to be located near their customers. But in a place near the state boundary, with another city just across the line, this matter becomes important. Many residents of New York City have thus avoided the city sales tax by transferring their custom to merchants in near-by places in New Jersey and Connecticut. And there is always the possibility of changing from the local merchant to the mail-order store.

Location of population. Taxation also influences the location of population. Anyone who has the choice of the town in which he will own a home gives thought to the local property tax rate, as well as other town or city taxes. It is well known that certain people choose Connecticut for their residence rather than New York, because the latter state imposes a personal income tax, while the former does not. The relative severity of the estate and inheritance taxes of the several states likewise affects the choice of residence of some persons who have retired from active business or are otherwise free to select their state of residence.

Form of tax. The form of a tax, as well as its mere presence, may have its own peculiar collateral effects. For example, under the United States federal income tax, a corporation, in computing its taxable net income, may deduct interest paid on bonds, but not dividends paid to owners of stock. This provision has a certain effect of encouraging the financing of business by borrowing rather than by the issue of capital stock.

Extravagant business expenditure. Heavy taxation of the net income or profits of business enterprise has the effect of stimulating business expenditures beyond what would otherwise be dictated by sound judgment. In the days of the federal excess profits tax of the Second World War, many a business corporation was in the posi-

tion of having to pay a tax of 80 per cent upon the last increment of its net income. Suppose the management of such a corporation were considering the pros and cons of an advertising campaign that would cost $100,000. If the expenditure is incurred, $100,000 will be deducted as a business cost, making the net income less by that amount. But if the expenditure is not incurred, the tax will take $80,000 of the $100,000 thus saved. The company may thus spend $100,000 at a cost to it, after taxes, of only $20,000. As was said, "Uncle Sam pays four-fifths of the cost." Often an expenditure that would not be considered worth its price may well be worth one-fifth of its price. Thus extravagant and wasteful expenditure by business enterprise is encouraged.

Incidentally, heavy business taxation of this sort defeats its own fiscal purpose by reducing the base upon which the tax is imposed. Business earnings tend to go into extravagant costs instead of income tax payments.

Charitable contributions. Severe taxes in the United States in recent years have had a curious double effect upon charitable contributions. The immediate effect of heavy taxes on incomes and estates has obviously been to diminish the funds out of which generous people, especially the more wealthy, make their contributions. On the other hand, a man who finds himself, let us say, in the 75 per cent bracket as to his last increment of net income, may receive the credit for a charitable contribution of $10,000 (deductible in computing taxable net income) at an actual cost to him of only $2,500. Solicitors for many a hospital or other charity have made much of this argument in recent years, to the great advantage of their institutions.

Effect on utilization of land. While taxation has in general a retarding effect upon industry, it may sometimes be true, as claimed by the apostles of the single tax, that a tax on economic rent or on land values has a stimulating effect upon the use of land. The owner who would otherwise hold his land idle will be forced by a tax, so the argument goes, to put his land to use.

It is true that the loss he suffers by holding his land idle is precisely the same, whether there is a tax or not. Suppose the land is capable of yielding a rental of $1,000 a year, and suppose the tax would amount to $300 a year. If the land is taxed, the owner's choice is between receiving $700 net if he rents his land, or paying $300 if he holds his land idle; holding the land idle means a net

sacrifice of $1,000. But if the land were not taxed, the difference between using the land and holding it idle would obviously be the rental of $1,000. The presence of the tax would have no influence upon the policy of one who followed his interests strictly.

On the other hand, it is probable that not every landowner calculates so meticulously as this. Paying out money for no return may be regarded more seriously than merely neglecting to get money that could be obtained. And there may be some who could forego a possible income, while not being financially in position to incur an avoidable expense. To this extent we may grant the stimulating effect of a tax on land.

Effect on industry: "Incentive Taxation." Barring this effect of land taxation there is no such thing as "incentive taxation," except in the sense of the schoolboy's essay on pins, in which he announced that "pins have saved a great many people's lives—by their not swallowing them." It is not taxation, but the escape from taxation, that can furnish an incentive to industry. One's business activity may be stimulated by a tax which is laid upon his competitors but not on himself. The tariff on imported cutlery is an incentive to the American manufacturer of pocketknives, who pays no such tax, as the production of butter is no doubt somewhat stimulated by the internal revenue tax on oleomargarine.

Reducing the chance of profit. As we learned in an earlier chapter, it is the hope of profits that induces entrepreneurs to assume the guidance and the risks of business. We have noted also the facts of actual profits and have seen that, in the long run, the net total of profits and losses together is a very modest sum.

Under a free economy, the existing prospect of profits is sufficient inducement to lead men to perform the entrepreneurial function of planning, organizing, and risk-bearing. By the same token, it is no more than sufficient; competition sees to that. If now we serve notice on the entrepreneurs that if their enterprises fail, they will bear the losses as usual, whereas if they succeed, heavy taxation will take 50 per cent, 60 per cent, 80 per cent of the profits, then the odds which were formerly sufficient to lead them on will, for some of them, no longer be good enough.

American investors in recent years, after carefully checking the prospects, have frequently decided not to put their capital into new business ventures, with the taxation cards thus stacked against them. With the odds thus rigged, many a businessman has decided

against carrying out his plans for adding a new wing to the factory or installing new machinery. Instead he has turned to the refuge of bonds, which yield him their comparatively small income, at a low rate indeed, but without the cares and labor and risk of business.

Production of the goods that satisfy human wants requires, in a free economy, the investment of capital and the performance of the entrepreneurs. Heavy taxation retards investment and enterprise and lowers the living standard of all the people.

Small family businesses. We have noted the great number of small independent businesses—individual proprietorships, partnerships, and small corporations—which, in spite of the trend toward large-scale enterprise, still exist and flourish in the American economy. Many a business of this class was built up from the start by a man of energy, skill, and daring. During his lifetime, he established a business and founded a modest family fortune to pass on to his descendants, making on the way his contribution to the supply of goods that satisfy human beings. Occasionally such a business (that of the late Henry Ford is an outstanding example) grows to colossal size. But most of these firms are small and inconspicuous, though none the less essential to the economic well-being of the community.

Under a tax regime of such severity as has prevailed in America of late, the opportunity to build up a business from small beginnings by "plowing back" the current profits is not what it was. Taxation each year takes a large share of the profits which could formerly have gone to increasing the business capital. It is hard to see how another Henry Ford could ever emerge under present conditions. What is of far greater importance, the future is much less promising now for all the thousands of young men who might like to build up their own businesses, following the example of their fathers.

Moderate vs. heavy taxation. It is hardly necessary to point out that the importance of all these collateral effects of taxation is directly related to the severity of the taxes. Under moderate taxation, some of these consequences would be of only trifling importance, and none would be really serious. Under very heavy taxation, some of them become a real threat to the public welfare. It is, for example, not taxation in general, but extremely heavy taxation, that threatens to slow down production and lower the standard of living.

Burden of compliance. Taxation is a deterrent to enterprise not only on account of the money actually collected. There is also the expense of compliance with the law. Some of the tax laws are extremely complicated. The parts of the United States *Internal Revenue Code* of October, 1944, covering income tax, excess profits tax, gift tax, and certain other tax provisions, make up a document of some 330 closely printed pages. The official *Regulations*, relating to income tax alone, constitute a volume of some 500 good-sized pages. Only an expert, or a corps of experts, can understand all its provisions and apply them to particular cases. Yet every taxpayer must know the law or suffer the consequence. The taxpayer must either have a staff of lawyers, accountants, etc., in his own organization or employ outside counsel. Voluminous and detailed forms must be filled out, requiring much time of the accounting and clerical staff of every business concern, not to mention time-consuming attention by top executives. All this represents a burden of taxation that is beyond the payment of the actual tax due and is commonly overlooked in popular judgments of taxation.

Stability of law and administration. The deterrent effect of taxation upon industry is greatly increased when tax laws are subject to frequent change, or when administrative officers have discretionary authority to modify important features of the laws. The businessman has to make his plans for the more or less distant future. Especially must he estimate the costs of future production. Since taxation today is one of the principal costs of business, uncertainty as to what his future tax liability may be is a serious handicap to the businessman. This is something that is overlooked by those who would delegate to an administrative body the discretionary power to change tax rates, in the effort to control the business cycle or for other purposes.

Nonfiscal taxation: Sumptuary taxes. Sometimes the nonfiscal effects become the primary purpose of taxation. There are thus the "sumptuary taxes," whose purpose is to control certain activities or the consumption of certain products that are regarded as contrary to the public interest. This idea is without doubt one of the forces that has led to the almost universal heavy taxation of alcoholic liquors and tobacco, and the license taxes imposed upon the manufacturers and sellers of these products. The long-standing federal excise tax on playing cards is in some measure tinged with this

purpose, though there is reason to doubt its effectiveness in curbing the great American game.

In most cases the sumptuary tax produces both revenue and, more or less successfully, the other ulterior effect. In general the two objectives are conflicting. Obviously the more successful a tax is in reducing the consumption of alcoholic liquor, the less revenue it will yield. In the case of the 10 per cent tax upon the notes of state banks, imposed by the United States at the time of the Civil War, the primary purpose was to put a stop to the issue of the state bank notes. So successful was the tax in this respect that it never produced any revenue.

"Protection." The protective tariff is in some ways similar to the sumptuary taxes. The customs tariff generally has the double purpose of raising revenue and giving protection to domestic industries. As with the sumptuary taxes, the two purposes war against each other. Protection results from keeping the taxed foreign goods out of the country. Revenue comes only from the foreign goods that enter. Complete success in giving protection would mean no revenue at all; the most effective revenue tariff would be one that gave no protection.

Sometimes a domestic industry is granted protection against another domestic industry, in this case by an internal excise tax. The United States excise taxes on oleomargarine, filled cheese, and renovated butter are for the purpose of increasing the selling prices of these products and so curtailing the strength of their competition against the butter and cheese industries.

Law by indirection. Taxation is sometimes employed to accomplish indirectly certain purposes which the government, from lack of constitutional authority or some other reason, is not disposed to attempt by direct legislation. Recently the United States Congress undertook the regulation of child labor in industrial employment. This is a matter over which the national government does not have authority under the Constitution. Congress therefore enacted a code of child labor regulations and then levied a tax upon any product entering interstate commerce which was produced in an establishment that did not conform to the code. This tax was held unconstitutional by the Supreme Court.

By a similar device, Congress attempted in 1935 to impose a code upon the bituminous coal industry. A tax was levied upon all

coal produced in the United States, at the rate of 15 per cent of the sale price at the mine, but the law authorized the remission of 90 per cent of the tax to producers who accepted and lived up to the code. This law also was held unconstitutional, but two years later another act was passed containing the substance of the previous act, with a similar tax feature. This act was carefully drawn with reference to the Supreme Court's opinion of the earlier act, and thus far it has not been found unconstitutional.

Controlling the business cycle. In recent years it has been proposed to use taxation as a means of regulating the industry of the country and the spending and saving of the people, especially for the purpose of controlling the business cycle. We have already noted the disturbing effect that such a program might have on business, due to arbitrary and unpredictable changes in tax rates and other tax provisions. Since these proposals involve more than taxation, including especially control of government spending, the budget, and public borrowing, this topic will be more conveniently discussed at another place.[1]

Arbitrary extremes. Going to the extreme of nonfiscal use of taxation, there is almost no limit to the regulatory, or even destructive, effects that may be accomplished by taxation in the hands of a totalitarian government or one in the control of a certain class seeking the domination or destruction of other classes. As the great Justice Marshall said in the year 1819 in delivering the opinion of the United States Supreme Court in the famous case of *McCulloch* v. *Maryland,* "the power to tax involves the power to destroy." Taxation may be used to impoverish certain persons and enrich others. It may foster or kill whole industries and destroy the prosperity of entire regions. It may direct the course of industry and trade, affect the movement and location of population, interfere with the distribution of wealth and income.

Having these considerations very sharply in mind, the framers of the Constitution of the United States, while undertaking to confer upon the federal government sufficient taxing powers to permit it to exist and function as a truly sovereign national government, took pains to impose strict limits upon the powers so granted. These have been mentioned at an earlier point in this chapter. In the years that have followed, the federal government has obtained a considerable widening of its powers of taxation, through amend-

[1] See Chapter xxvi.

ment of the Constitution and more especially through court decisions interpreting more and more liberally the constitutional limits upon its powers. Probably the most far-reaching of all steps in this direction was the adoption of the Sixteenth Amendment to the Constitution, which gave the federal government power to levy an income tax without complying with the requirement that direct taxes must be apportioned among the states according to their population.

CHAPTER XXVI

American Public Finance

The cost of government: Public expenditures and national income. In the year 1940 there were expended by government in the United States about 18¼ billions of dollars. We select 1940 as the last year before Second World War activities began to dominate the national government's finances. The federal government's share was about 9 billions; the forty-eight states spent around 3⅔ billions; and the various local governments in all the states accounted for 5⅔ billions. These are huge sums, beyond the power of the imagination to visualize. Somewhat easier to grasp is the fact that government expenditure in that year took nearly a quarter of the total income of the people of the United States, estimated at 77⅔ billion dollars.

If perchance the situation in 1940 may be considered abnormal, on account of extraordinary expenditures for social purposes by the federal government during the depression decade, we may go back to the year 1929, when the total expenditures of government were about 11⅔ billions and represented 14 per cent of a national income of 83⅓ billions. It will be noted that between these dates there was both a decline in the people's income and a heavy increase in the cost of government.

The record of public expenditure: General summary. The constant growth of public expenditures is an outstanding phenomenon of modern history. Starting as far back as reliable records will take us, this universal tendency appears. It applies to all types and grades of government: monarchies and republics, peaceful nations like Switzerland as well as militaristic Germany, imperial governments, national governments, states, and municipalities.

The increase in government expenditures before the nineteenth century, however, is on the whole more apparent than real, the great advances in the absolute figures being substantially offset by such factors as diminishing reliance on goods and services directly rendered, growth of population, increase in territory, increased

wealth of the people, and depreciation of the monetary standard. But since the beginning of the nineteenth century, there has undoubtedly been real growth.

Federal government of the United States. The two tables on this page and the next page present a historical record of public expenditures in the United States of America.

For the first seventy years of the national existence of the United States, the expenditures of its government were on a modest scale,

GOVERNMENT EXPENDITURES IN THE UNITED STATES[1]
(In millions of dollars)

Year	Federal	State	Local	Total
1792	5			
1800	11			
1810	8			
1820	18			
1830	15			
1840	24			
1850	40			
1860	63			
1870	310			
1880	268			
1890	318	77	487	882
1900	521			
1903	517	182	913	1,612
1910	694			
1915	761	483	2,073	3,317
1919	18,515	625	3,041	22,181
1920	6,403	789	4,035	11,227
1925	3,063	1,479	5,196	9,738
1930	3,440	2,152	6,071	11,663
1935	6,592	2,396	4,583	13,571
1940	9,305	3,612	5,640	18,557
1941	13,766	3,585	5,553	22,904
1942	34,289	3,644	5,219	43,152
1943	79,702	3,503	4,873	88,078
1944	95,572	3,588	4,604	103,764
1945	100,397	3,797	5,015	109,209
1946	63,714	4,416	5,121	73,251
1947	42,505	4,000	5,000	51,505
1948	36,326			

[1] The figures of expenditures of the federal government through 1947 are from the *Annual Report of the Secretary of the Treasury on the State of the Finances, 1947*, pp. 272–77; the 1948 figure is from the *Daily Statement of the United States Treasury*, June 30, 1948. The state and local figures through 1946 are from the National Industrial Conference Board, *Cost of Government in the United States, 1927–28*, New York, 1930, p. 2; and the board's *Economic Almanac, 1946–47*, and *Economic Almanac, 1948*. The state and local figures for 1947 are based on estimates by the Finance Department of the Chamber of Commerce of the United States. The net expenditures of the wholly owned government corporations are included in the expenditures of the federal government for 1935 through 1948.

the increase from 11 millions in 1800 to 63 millions in 1860 being no more than sufficient to keep pace with the growth of the country's population, as is shown by the per capita expenditures.

The figures for 1870 show the effect of the Civil War cost, and the later figures illustrate the universal truth that war not only adds a wartime burden of extraordinary costs, but sets a new scale of peacetime expenditure from which no people has ever been able to

GOVERNMENT EXPENDITURES IN THE UNITED STATES[1]

(Dollars per capita)

Year	Federal	State	Local	Total
1792	1.28			
1800	2.03			
1810	1.13			
1820	1.89			
1830	1.18			
1840	1.42			
1850	1.71			
1860	2.01			
1870	8.03			
1880	5.34			
1890	5.05	1.22	7.73	14.00
1900	6.85	——	——	——
1903	6.40	2.25	11.27	19.92
1910	7.54	——	——	——
1915	7.66	4.80	20.61	33.07
1919	177.15	5.97	29.09	212.21
1920	60.57	7.46	38.17	106.20
1925	26.44	12.77	44.85	84.06
1930	27.93	17.52	49.45	94.90
1935	51.90	18.83	36.02	106.75
1940	70.49	27.37	42.65	140.51
1941	103.50	26.91	41.69	172.10
1942	253.99	27.04	38.67	319.70
1943	583.90	25.67	35.37	644.94
1944	692.55	25.98	33.34	751.87
1945	719.18	27.20	35.92	782.30
1946	451.87	31.32	36.32	519.51
1947	297.24	27.97	34.97	360.18

return to the prewar status. Other influences contributed further to the growth of public expenditures, so that, from the Civil War to the First World War, the cost of the federal government averaged more than four times as much as before the Civil War, after making full allowance for the growth of population.

[1] Except where given in the sources, the figures in this table have been computed by dividing the figures of the preceding table by the mid-year estimates made by the Bureau of the Census of population of the United States.

The gigantic expenditures of the First World War reached their maximum in 1919. Expenditures then dropped, but they were never again as low as 3 billions. The new level reached in 1925 was four times as high as the expenditures of 1915, or three and a half times as high after allowing for the increase in population.

The next major rise in the level of federal expenditures began in 1932, chiefly the result of financial measures taken with the double purpose of counteracting the depression that followed the crisis of 1929 and bringing about greater equality in the distribution of the national wealth and income. Governmental policy during the decade of the 1930's was influenced by doctrines which inclined to consider public expenditure desirable in itself, as a means of "putting money in circulation," stimulating purchasing power, "priming the pump," and restoring normal industrial activity. From this viewpoint, all government spending carried a favorable presumption, tending to brush aside traditional notions of economy and the balanced budget.

As a result, the expenditures of the national government rose rapidly, as shown by the figures for 1930 and 1940 in our tables. By 1940 the cost of the federal government was nearly three times what it was ten years before.

From here on war takes the center of the stage. Federal expenditures in 1945—mostly for war purposes, of course—reached their unprecedented maximum of 100 billion dollars.

Billions of dollars are pretty hard to visualize. The magnitude of the government's expenditures in this period may perhaps be made more realistic by considering some comparisons like these: The total cost of the federal government in each of the years 1944 and 1945 was more than five times the cost of the highest year in the First World War (18½ billion dollars in 1919). The cost in each of the years 1944 and 1945 equaled the total national income in the year 1942. Payment for interest alone in 1944 and in 1945 was more than the total cost of the federal government in any year from the First World War to the depression. In the one year, 1945, the government spent more than in the whole nineteen-year period 1922–40. The per capita expenditure in 1945 was twenty-four times what it was ($30) in the decade of the 1920's.

Following the close of hostilities in the summer of 1945, war costs naturally declined, with corresponding reduction in the cost of

government in 1946 and 1947. Even so, military costs for defense were still the major feature.

American state and local governments. It is only since the latter part of the nineteenth century that we have data enabling us to bring state and local expenditures into the picture of the growth of all government costs in the United States. It will be observed from the tables on pages 509 and 510 that the expenditures of the states show a fairly consistent record of rapid increase from 1890 to 1940, after which they stabilized in the neighborhood of 3 to 4 billion dollars. Local expenditures likewise staged a rapid increase beginning in 1890, but reached a state of comparative equilibrium some twenty years earlier than the state governments.

In explanation of the contrasting records of the federal government and the other two grades, we note (1) that the latter do not have the responsibility for war and defense, (2) that they have not been so much influenced by doctrines of public spending for control of the business cycle or redistribution of wealth, and (3) that they are more limited as to their revenue resources; this latter consideration applies especially to the nonsovereign local bodies.

Distribution between federal, state, and local. In the year 1890, the total cost of government in America was distributed thus: federal government, 36 per cent; state governments, 9 per cent; local governments, 55 per cent. It is interesting to note that the three grades of government ran very nearly abreast in the race for increased expenditures up to the "great depression." In 1929, the distribution was: federal, 28 per cent; states, 16 per cent; local, 56 per cent. The states had gained somewhat, at the expense of the federal government.

From that time on, as we could infer from the absolute figures, the federal government has been steadily gaining at the expense of the two other grades. In 1940 the federal government accounted for 49 per cent of the total cost of government, and in 1945, at the climax of war expenditures, its share was 92 per cent. Two years later the federal government was, according to the estimates, still accountable for 83 per cent of the total.

Functional analysis. One naturally raises the question: What does the government spend its money for? Figure 32 presents a summary answer in the form of a functional analysis of the expenditures of each grade of government in the year 1941.

These figures are intended to show the net expenditures of each

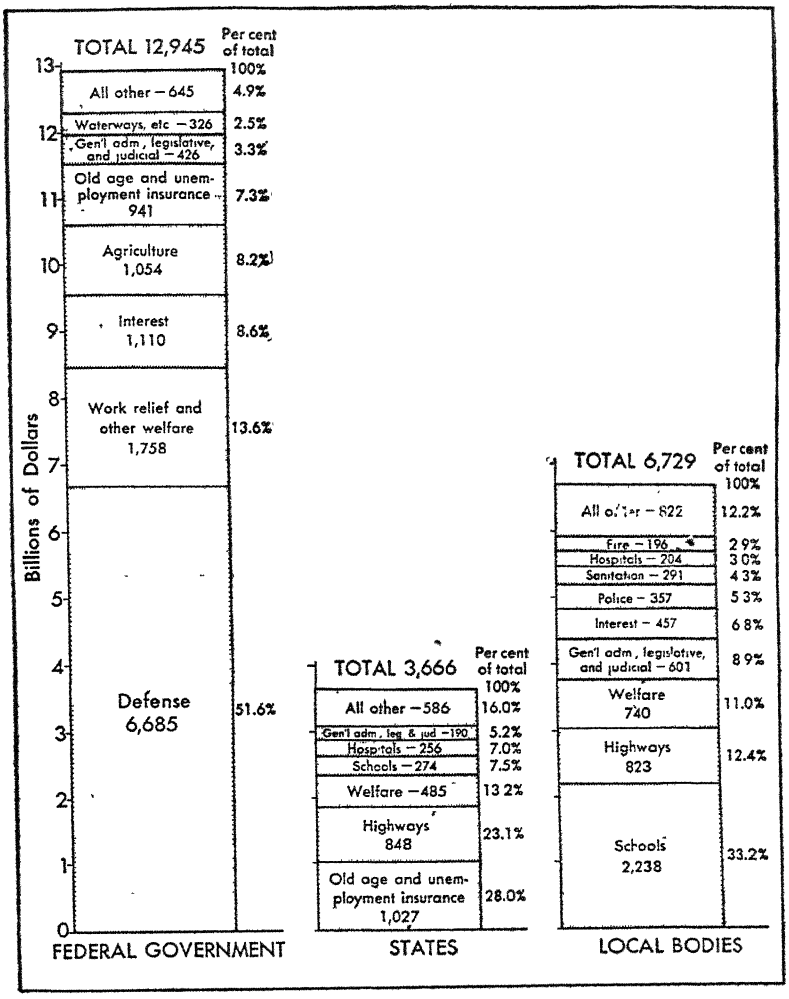


FIG. 32. EXPENDITURES OF AMERICAN GOVERNMENTS, BY FUNCTIONS, 1941.[1]

[1] Adapted from Bureau of the Census, *Financing Federal, State, and Local Governments, 1941*. Washington: Government Printing Office, Sept., 1942. Table 17, p. 54.

grade of government for its own functions. They therefore exclude payments by any unit of government to other public units, but they include payments made from funds received from other public units. For example, when a state contributes money to a town to be spent by it, this sum is not included in the expenditures of the state; it is added to the money received by the town from its own revenues and is included in the analysis of the town's expenditures for its own functions.

This chart speaks for itself. Starting with the federal government, it is obvious that defense is the principal function as measured by cost, accounting for more than half of the total. In 1941, as it happens, the cost of the Second World War was already swelling this item. But in any other year, even though peace prevails, the costs of past wars and of defense or provision for future wars stand first. This is generally true of all national governments.

It will be observed that the next three items in order of cost are for work relief and other welfare, interest, and agriculture, and that all the general administrative, legislative, and judicial business of the government required only a little over 3 per cent of the total money spent.

Old age and unemployment insurance stands first among the objects of expenditure of the American states, accounting for more than one-fourth of their total expenditures. Next in order come highways and the various forms of welfare. Schools and hospitals are next, with about equal costs.

The principal item of cost among the local governments is for schools, accounting for about one-third of their total costs. Highways come second, and welfare third.

The revenue system. The term "revenue," in the language of public finance, includes all receipts of a government that add to its net assets. This excludes such receipts as the proceeds of loans or money received for trust funds that do not belong to the government. The great bulk of the revenues of all American governments is made up of taxes, and our present discussion is concerned chiefly with taxation.

Federal government. We have noted that among the constitutional limitations upon the national government's taxing power is the provision that direct taxation (other than that of incomes as provided by the Sixteenth Amendment) may be employed only if apportioned among the states according to their population. This

rule made the practical task of levying a direct tax almost prohibitively difficult. Some timid experiments were made soon after the adoption of the Constitution and again at the time of the Civil War. But these direct taxes were ineffective and produced little revenue, and the attempt has never been repeated.

From the start the national government relied upon indirect taxation. Of the two principal types of indirect taxes, the customs tariff was at the first chosen rather than the more obvious internal excise taxes, again on account of popular antagonism to taxation. While there was some half-hearted use of excise taxes up to the close of the War of 1812, the tariff was practically the only tax resource of the national government until the time of the Civil War. It proved, as we have learned, an erratic source of revenue at best, and it failed completely to meet the emergency needs of war.

The Civil War was of such duration as to give the people time to overcome their traditional repugnance to internal taxation. Starting in 1862, excise taxes were enacted, which finally laid burdens upon almost every article of industry and commerce, and in the financial reorganization after that war, the excise system, though greatly reduced by freeing the majority of taxed commodities, was not entirely abandoned as it had been after the War of 1812. A group of excises, principally upon liquors and tobacco, remained and provided, not only an important source of peacetime revenue, but also the skeleton upon which it was possible to build quickly an effective system of war taxes in time of future need. Good use was made of this on the occasions of the Spanish War and the two World Wars of the twentieth century.

For the half century following the close of the Civil War, the tariff and the excises shared about equally (tariff revenue being generally a little greater than excise revenue) in the task of producing the revenue needed by the federal government. No other form of taxation, and indeed no other source of revenue of any sort, was of comparable significance. For example, during the generation ending in 1913, the combined receipts from customs and excises provided more than 90 per cent of the net ordinary revenue (*i.e.*, excluding loans) of the national government in every year except one.[1]

[1] In 1898 this percentage was reduced on account of a large fortuitous income from sale of public property to the Kansas Pacific Railroad and the Union Pacific Railroad.

The federal tax system was revolutionized by the permanent addition of the income tax—on corporations in 1909 and on individuals in 1913. Shortly thereafter the estate tax, regarded heretofore as a temporary war measure, became a fixed part of the system when Congress failed to repeal the estate tax of the First World War.

In the fiscal year 1929 *(i.e.,* the twelve months ending June 30) the federal government obtained its revenues ("ordinary receipts") as follows:

REVENUES OF THE UNITED STATES GOVERNMENT, 1929[1]

	In millions of dollars	Per cent
Customs	602	14.9
Internal revenue (miscl.)	546	13.5
Income and tax profits	2,331	57.8
Estate tax	62	1.5
Other ordinary receipts	493	12.2
Total ordinary receipts	4,033	100.0

Nothing could show more clearly the revolution in the revenue system of the federal government. Up to 1913, as we have seen, the national government had relied almost exclusively upon indirect taxation, leaving the whole field of direct taxes to the states and their local subdivisions. Thereafter the federal government not only invaded the direct tax field, but it made direct taxation its chief reliance. In 1929 more than half of the total ordinary receipts came from the income tax. Barely over one-fourth came from the two great indirect taxes, customs and excise, which for a generation before 1913 produced more than 90 per cent of the ordinary receipts.

The depression following 1929 played havoc with the federal revenue system. The yield of the income and excess profits taxes, upon which chief reliance then rested, suffered a drastic decline; in 1933 they produced only 746 millions, a loss of two-thirds of their yield since 1929. Total revenue receipts in 1932 and 1933 were only half what they were in 1929.

Revenue legislation was enacted in each of the next six years. These acts increased heavily the rates of the income tax, both indi-

[1] *Annual Report of the Secretary of the Treasury on the State of the Finances, 1947.* Washington: Government Printing Office, 1948, pp. 276, 311.

vidual and corporate, and the estate tax, and by 1938 the yield of these taxes had been brought back about to the level of 1929.

It was the excise taxes, however, that really came to the aid of the Treasury. The legislation of 1934–39 broadened the base of excise taxation and raised rates. By 1938 their yield had quadrupled, and their percentage contribution to the total revenue was approaching the position of the years before the income tax appeared.

These developments are apparent in the following figures for 1938.

REVENUES OF THE UNITED STATES GOVERNMENT, 1938[1]

	In millions of dollars	Per cent
Customs	359	6.1
Internal revenue (miscl.)	2,230	38.1
Income and profits tax	2,640	45.1
Estate and gift tax	417	7.1
Other ordinary receipts	208	3.6
Total ordinary receipts	5,855	100.0

The Second World War (1941–45) imposed unprecedented demands upon the federal revenue system. Without going into the details of war tax legislation, it is to be noted that no additional contribution was obtained from the customs tariff, and only a little from estate and gift taxes. The excise system, on the other hand, was called upon for a heavy contribution. But the chief demand was upon the income and profits taxes. Rates were raised all along the line, and the yield was prodigious. Once more the center of gravity of the tax system was far on the side of the direct taxes.

Revenue receipts were at their maximum in the fiscal year 1945, when they totaled 46,457 million dollars. Figure 33 on page 518 gives a picture of the principal taxes and their relative position in the revenue system of that year.

It will be observed that, although the excises had nearly trebled in absolute amount, their share of the total had droppd from 38 per cent in 1938 to 13 per cent in 1945. Income and profits taxes, on the other hand, yielding the unprecedented sum of 35 billion dollars, contributed three-fourths of the total revenue.

The states. In the early years of the American nation, the states relied for revenue principally upon the property tax. As time went

[1] *Ibid.*

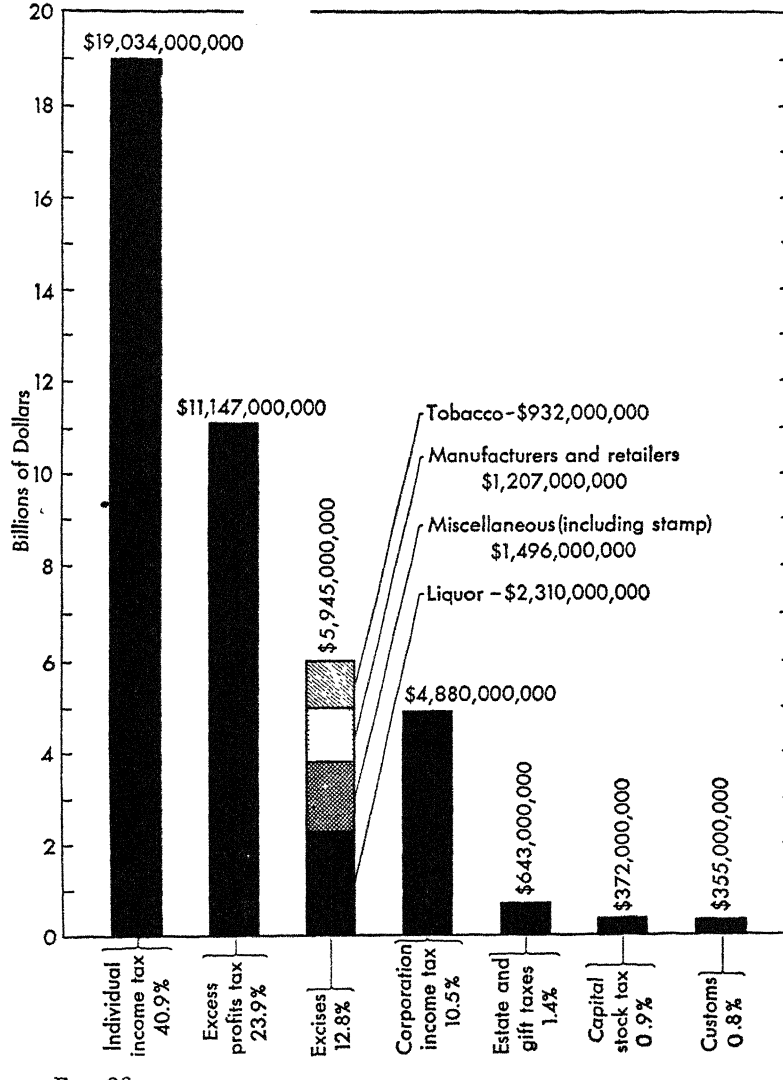


FIG. 33. TAX RECEIPTS OF THE UNITED STATES GOVERNMENT, BY SOURCES, 1945.[1]

[1] Figures from *Annual Report of the Secretary of the Treasury on the State of the Finances, 1945*. Washington: Government Printing Office, 1946.

REVENUES OF THE AMERICAN STATES, 1946[1]

	Number of states	Millions of dollars	Per cent of total taxes excluding unemployment compensation
Sales and use taxes			
General	23	900	18.1
Motor vehicle fuel	48	887	17.8
Alcoholic beverages	48	402	8.1
Tobacco products	31	199	4.0
Pari-mutuel	20	103	2.1
Admissions and amusements	29	12	0.2
Other	30	26	0.5
Total sales and use taxes		2,529	50.8
Gross receipts taxes			
Insurance companies	48	145	2.9
Public utilities	37	152	3.1
Total gross receipts taxes		297	6.0
License and privilege taxes			
Motor vehicles	48	429	8.6
Motor vehicle operators	48	33	0.6
Corporations in general	47	120	2.4
Alcoholic beverages	48	67	1.3
Hunting and fishing	48	32	0.6
Occupations	48	14	0.3
Chain stores	20	4	0.1
Amusements, race tracks	31	4	0.1
Other	48	84	1.7
Total license and privilege taxes		787	15.8
Individual income tax	33	389	7.8
Corporation income tax	33	442	8.9
Property tax	46	253	5·1
Death and gift taxes	47	146	2.9
Severance tax	23	90	1.8
Poll tax	10	7	0.1
Documentary and stock transfer taxes	14	39	0.8
Total taxes (excl. unemployment compensation)		4,980	100.0
Unemployment compensation taxes	48	1,034	20.7
Total taxes (incl. unemployment compensation)		6,041	120.7
Nontax revenue	48	466	9.3
Total revenue from own sources		6,480	130.0
Aid received from other governments			
Federal grants		743	14.9
Local grants		24	0.5
Total revenue from all sources		7,246	145.4

[1] Bureau of the Census, *State Finances 1946*. Washington: Government Printing Office, 1947. Vol. 3, *Statistical Compendium*, Tables 5, 6, 7, and 8.

on, a gradual transition took place. The states step by step added other forms of taxation and reduced their demands upon the property tax, leaving the latter more and more to the local bodies. In 1946 there were two states that made no use of the property tax. A list of the principal taxes used by the states, with the amounts raised by each and their relative importance, is given in the table on page 519.

This table shows also the extent to which the states as a whole receive nontax revenues and assistance from other grades of government, chiefly federal aid.

Local governments. The following table summarizes the revenues of the several types of local government in the year 1941:

REVENUES OF AMERICAN LOCAL GOVERNMENTS, 1941[1]

(Thousands of dollars)

Type of revenue	County	Municipal	School districts	Special districts	Total
From own revenue collection					
Property tax	925,072	2,213,085	968,404	117,172	4,223,733
Other taxes	69,362	299,037	11,686	1,938	382,023
Special assessments	10,870	63,181		28,070	102,121
Other nontax revenue	145,008	304,949	34,472	59,948	544,377
Total own sources	1,150,312	2,880,252	1,014,562	207,128	5,252,254
Fiscal aid	626,810	560,018	587,955	18,881	1,793,664
Total revenues	1,777,122	3,440,270	1,602,517	226,009	7,045,918

Two significant features of the local revenue systems are (1) the comparatively heavy reliance on fiscal aid from other governments —chiefly state aid and federal aid—and (2) the predominance of the property tax among their own revenue sources.

Fiscal aid made up the following percentages of total revenues:

Counties	35.3 *per cent*
Municipalities	16.3
School districts	36.7
Special districts	8.4
All local governments	25.5

The preponderance of the property tax is shown by the per cent of own revenue collections which this tax constitutes, as appears in the table on the opposite page.

[1] Adapted from Bureau of the Census, *Financing Federal, State, and Local Governments: 1941.* Washington: Government Printing Office, September, 1942, p. 9.

Counties	80 *per cent*
Municipalities	77
School districts	95
Special districts	57
All local governments	81

It is only the special districts that obtain anything like half their own revenues from other sources. They obtain 13 per cent from special assessments, and 30 per cent from other nontax revenues.

Public borrowing: Purposes. A government borrows in order to counteract a deficit, realized or anticipated. The circumstances which cause the deficit and so compel the government to borrow are various.

In the absence of extraordinary events which may arise to affect either the expenditure or the income of a government, there is normally an approach toward agreement between the totals on the two sides of the account. Even in the best regulated governments, however, an exact balance is seldom to be looked for. Moderate deficits and surpluses occur from time to time, and while in the long run they may tend to offset each other, the periods of deficit require borrowing to make both ends meet.

Such borrowing in order to meet a deficit in the ordinary running expenses is common to all governments. Especially in the United States, where budgetary procedure and financial administration are far from perfect, federal, state, and local governments are frequent borrowers for this purpose.

No nation can meet the extraordinary expenditures of a great war out of its ordinary sources of revenue, however heroically it may expand its tax system and other revenues. Moreover the magnitude of war costs is seldom foreseen by the people and their leaders at the war's beginning; realization of the extraordinary contributions which will be required comes slowly, and the war is generally well along, sometimes approaching its close, before the maximum financial effort of which the people are capable is exerted. Often a weak or unintelligent leadership purposely avoids calling upon the people for heavy tax contributions.

For these reasons war generally causes large government loans. The American Revolution was almost entirely financed by loans. The cost of the Civil War was about 3 billion dollars, of which four-fifths was borrowed, the balance being met mainly by taxation. The First World War was the occasion of the most gigantic loans

in the history of public finance to that time. The United States borrowed about 24 billion dollars, or three-fourths of the cost imposed by the war. Great Britain expended 42 billion dollars on account of the war, obtaining 36 billions by means of borrowing. Germany, under the guidance of a weak and mistaken leadership, borrowed almost the entire amount of her war expenditure. Finally all records were again broken in the Second World War, when the United States, for example, added 230 billion dollars to its public debt between June 30, 1940, and May 31, 1946.

Local governments and members of federal governments, like the states of the United States, do not have to finance wars, and since they are seldom confronted with other serious crises, emergency borrowing is comparatively rare with them.

When a government engages in an industrial enterprise of any magnitude, it usually finds it necessary to borrow the capital required. When the Prussian government decided to take over the railway system, it borrowed the money to purchase the existing railroads from their private owners. The United States Government borrowed in order to construct the Panama Canal. The state of New York borrowed large sums to enlarge and improve the Erie Canal. American cities occasionally borrow in order to establish waterworks systems and other municipal industries. Governments also borrow not infrequently the money necessary to pay the cost of buildings and other public works needed for the performance of their primary functions. States and towns frequently borrow to build hospitals, asylums, and bridges; towns and cities, to build schoolhouses and other public buildings.

The security back of public debts. Government debts differ from ordinary private obligations in that there is usually no pledge of security to guarantee payment. The exception to this statement relates to weak governments with doubtful credit, which, when borrowing from foreign capitalists, are sometimes required to pledge certain public property or certain sources of revenue. Such pledges are, however, the exception, not the rule, and in the case of a strong nation no pledge whatever is ordinarily given beyond the word of the government. In case of default of the ordinary public debt, the creditor has no recourse except either to bring suit in the debtor government's own courts—which is not of much help if the government has deliberately chosen not to pay—or (if a foreigner) to seek

to persuade his own government to take diplomatic or, in the last resort, military action against the defaulting government.

High standing of government obligations. It may appear strange, in the face of the foregoing, that the unsecured debts of sovereign powers have generally enjoyed such high standing. As a matter of fact, government obligations, especially those of strong national governments, which are almost invariably unsecured, have been held in high favor by the investing public. Under normal conditions, governments could generally borrow at lower rates than were paid by other borrowers, and government obligations were generally quoted on the markets at lower yields than other investments. Bonds of respectable governments stood in the very front rank of "gilt-edged" securities.

The high standing of public obligations rests on certain characteristics of government. In the first place, the life of a government is not limited as is the life of an individual, and it is considered more enduring even than the lives of most corporations. Of chief significance, however, is the fact that the government's ability to meet its obligations, both interest and principal, is not dependent upon the success of any particular business venture. Most borrowing by individuals and corporations is for investment in business. If the business fails, the borrowers may default on the interest or on part or the whole of the principal. The government's ability to meet its obligations rests upon the taxing power, which may always be exerted to obtain the necessary money, provided the debt is not too large. The absence of security is counteracted by a most scrupulous regard for its financial obligations, by which governments have generally sought to protect their credit.

The fact is, however, that the history of the past generation has impaired the high standing of national government obligations. The First World War placed upon certain nations a financial strain which was more than they could bear. As a result of budget deficits and reckless issue of paper money, Russia, Germany, and certain other great nations virtually repudiated their entire public debts. Other nations saw their obligations sink far below par and suffered serious loss of credit; among these were France and Italy. Great Britain and the United States were the two conspicuous examples of great nations which came through that struggle without serious impairment of national credit.

Since then monetary depreciation has lowered the value of public debts in many nations. The action of the United States in 1933, in simultaneously repudiating the gold contract in its national bonds and depreciating the value of the paper money in which alone the bonds could thereafter be redeemed, was not favorable to the credit of the United States Government.

Then came the Second World War, imposing strains upon public credit heavier than ever before. Finally the very size of the debts of some of the leading nations—including the United States—is an element of weakness. Government obligations in only a few nations outside the United States today command that confidence which they held a generation ago.

The national debt of the United States. The War of the Revolution was financed principally by loans, and when the new national government of the United States started operations in 1789, it had a heritage of 52 millions of public debt—a sum which appears trivial enough to us of the present generation, accustomed to speak casually in terms of billions, but which the people of that day regarded as a heavy burden upon the shoulders of the new government.

Nevertheless one of the first acts of the government, under the far-sighted leadership of Alexander Hamilton, was the assumption, as an addition to the national debt, of the debts of the several states incurred in the prosecution of the Revolutionary War. This brought the debt up to 76 millions in 1794. American public opinion, from the very beginning, has looked with disfavor upon a large national debt and has always favored rapid payment. By 1812 the debt of the United States had been reduced to 45 millions. The War of 1812, likewise financed principally by loans, brought the national debt up to 128 millions.

Then commenced a period of rapid debt payment, culminating in 1836 in the complete retirement of the national debt, except for a few thousand dollars, which were unpaid because the holders had not presented their bonds, but for whose payment the Treasury had the cash in hand. This represents a unique performance in the records of public finance—a national debt completely paid off.

The country was not long out of debt. Deficiencies in current revenues and the Mexican War restored the debt, which in 1860, just before the Civil War, amounted to 65 million dollars. The borrowing of this war brought the national debt to 2,845 million dollars on August 31, 1865. Once more the nation set about the task

of debt payment and, by a performance only less remarkable than its previous record, had brought the debt in 1893 down to 839 millions, a reduction of more than two thirds.

During the next quarter century, there were numerous debt operations, increases as well as reductions, the principal causes of new borrowing being the resumption of specie payment of the greenbacks in 1879, the currency disorders of 1893–96, the financing of the Spanish War, and the digging of the Panama Canal. In 1916, the year before the United States entered the First World War, the national debt amounted to about a billion and a quarter. The First World War was financed to the extent of three-fourths of its cost to the United States by borrowing. The total borrowing on account of this war made everything preceding pale into relative insignificance; on August 31, 1919, the national debt amounted to 26½ billion dollars.

Once again the traditional American policy of debt reduction was called into operation and vigorously pursued. Large appropriations for payment of the debt were included in the regular annual budgets; in addition, substantial surplus revenues were regularly available for debt retirement. By December 31, 1930, the debt had been reduced to 16 billions, a reduction of two-fifths of the debt in eleven years, at the rate of nearly a billion dollars a year.

We have already noted the record of lavish public expenditure during the decade of the 1930's. Since revenues were not correspondingly increased (during the first years they actually declined), the national debt rose. By June 30, 1934, all the gain made since the war had been lost, and at the end of the fiscal year 1940 (June 30) the debt stood at 43 billion dollars, an increase of 27 billions in a decade of peace.

The enormous cost of the Second World War was, in spite of a vigorous tax policy, of necessity financed chiefly by borrowing. The national debt reached its maximum on May 31, 1946, when it stood at 272½ billion dollars.

The budget. Under sovereign constitutional governments, all taxes and other revenues and all expenditures must be authorized by law. This requires that at certain stated intervals the lawmaking body—be it Parliament, Congress, state legislature, or other lawmaking authority—must take account of the state of the public finances, determine what services shall be performed by the government, estimate the probable cost of such services, and provide taxes

and other income to meet the anticipated expenditures. In determining its fiscal policy, the legislature will find it helpful to take account of the income and expenditure of the current and immediately preceding years. Such account, together with the estimates and plans for the coming year, is called the *budget.*

The subordinate grades of government—counties, cities, towns, school districts, etc.—also have use for the budget. Although they have no sovereign power, they have been delegated a considerable authority over their own finances, under which the local legislative body—board of aldermen, city council, town selectmen, or whatever it may be—determines the expenditures and at least the rates of taxation for the district.

Budget procedure has been brought to a high degree of perfection by certain governments, the most notable being the British Parliament. In the United States, the budget has been generally more honored in the breach than in the observance by all grades of government. Only since the First World War has the federal government developed any semblance of a budgetary procedure, and barring a few notable exceptions, the states and municipalities have been no less backward.

Development of an efficient national budget has been hindered in the federal government by the constitutional separation between the executive and legislative arms of government and the bicameral Congress, features which go a long way to explain the inferiority of American budget practice to that of the British.

One of the chief weaknesses of the Congressional budgetary procedure was the long-standing practice of having revenue legislation and appropriations separately handled. The former was considered by the Senate Finance Committee and the Ways and Means Committee of the House, the latter by various appropriations committees in both branches of Congress. After receiving reports of committees, Congress enacted appropriation acts and revenue acts at various times during its session, and with little relation to each other.

In the British House of Commons, on the other hand, revenues and expenditures are both considered by a committee of the whole house, and a budget law is finally enacted which both authorizes all expenditures and imposes the necessary taxes or other revenue measures. Obviously the American practice lacked the most essential feature of a real budget.

A long forward step was taken in the United States in the Legislative Reorganization Act of 1946, which, among other things, provided for joint meetings of the taxing and appropriating committees of both Senate and House, from which should emerge a "legislative budget for the ensuing fiscal year, including the estimated over-all Federal receipts and expenditures for such year."

The balance of revenues and expenditures. Where revenues and expenditures are out of balance, there results either a surplus or a deficit, reflected eventually in a decrease or an increase of the public debt. As we have observed, the traditional American attitude has strongly favored the "balanced budget", and by and large this aim was accomplished throughout the history of the United States up to the beginning of the First World War. At that time the national debt amounted to barely over a billion dollars. This means that up to that time ordinary deficits had been matched by surpluses and even the borrowings of war had been mostly paid off.

The First World War produced the inevitable deficits of war, whose magnitude, as we have seen, is reflected in the increase of the public debt. Thereafter came an eleven-year period (1920–30) of surplus revenues, averaging nearly a billion dollars a year.

The next sixteen years present a record of continuous and generally large deficits. The record, including the last two years in which prosperity still smiled upon the national treasury, is summarized in the table on the following page and in Figure 34 on page 529.

The reader is warned against gaining a possible false impression from Figure 34. So high are the blocks portraying revenues and costs of the war years that the blocks for the decade preceding may look modest indeed. This would be an illusion. Actually the expenditures in the 1930's were unprecedentedly high for a period of peace, double the amount it had cost to run the government in the preceding decade.

Two principal causes are accountable for the deficits of the 1930's. In the first place, economic depression had its effect upon the revenues, especially the yield of the income tax. This is clearly discernible in the decline of receipts in 1931, 1932, and 1933, bringing the revenue down to only about half of the amounts received in 1929 and 1930. The revenue legislation of the next five years, together with improvement in general economic conditions, brought about a progressive increase in revenues through the fiscal year

UNITED STATES GOVERNMENT,
REVENUES AND EXPENDITURES, 1929–48[1]

(Millions of dollars)

Year	Revenue receipts	Cost payments	Surplus (+) or deficit (−)
1929	4,033	3,299	+734
1930	4,178	3,440	+738
1931	3,190	3,652	−462
1932	2,006	4,741	−2,735
1933	2,080	4,681	−2,602
1934	3,116	6,745	−3,630
1935	3,800	6,592	−2,791
1936	4,116	8,541	−4,425
1937	5,029	7,806	−2,777
1938	5,855	7,031	−1,177
1939	5,165	9,027	−3,862
1940	5,387	9,305	−3,918
1941	7,607	13,766	−6,159
1942	12,799	34,289	−21,490
1943	22,282	79,702	−57,420
1944	44,149	95,572	−51,423
1945	46,457	100,397	−53,941
1946	43,038	63,714	−20,676
1947	43,259	42,505	+753
1948	44,746	36,326	+8,419

Note. In this table and in Figure 34 trust fund transactions are, properly, excluded from the figures for the years from 1931 on. Owing to the form of the official reports, it is not possible to exclude these transactions from the figures for 1929 and 1930; The differences are not sufficient to affect seriously the record here presented.

1938. The decline in revenues thereafter reflects the depressed condition of national industry which prevailed from the fall of 1937 to the middle of the calendar year 1939.

At the same time, efforts to fight economic depression with public spending, starting in 1931, were getting into full swing. It will be noted that expenditures rose to over 8½ billions in 1936 and then declined slightly in the next two years. The heavy increases for 1939, 1940, and 1941 were consequent upon the pump-priming policy employed to meet the return of depression conditions in the fall of 1937.

As a consequence of these two causes, we have a decade of unbalanced finance, the deficit averaging 2¾ billions a year and continuing up to the beginning of the Second World War. The fiscal year 1938 came nearer to a balance than any previous year since

[1] *Annual Report of the Secretary of the Treasury on the State of the Finances, 1947.* Washington: Government Printing Office, 1948, pp. 276–7; *Daily Statement of the United States Treasury,* June 30, 1948.

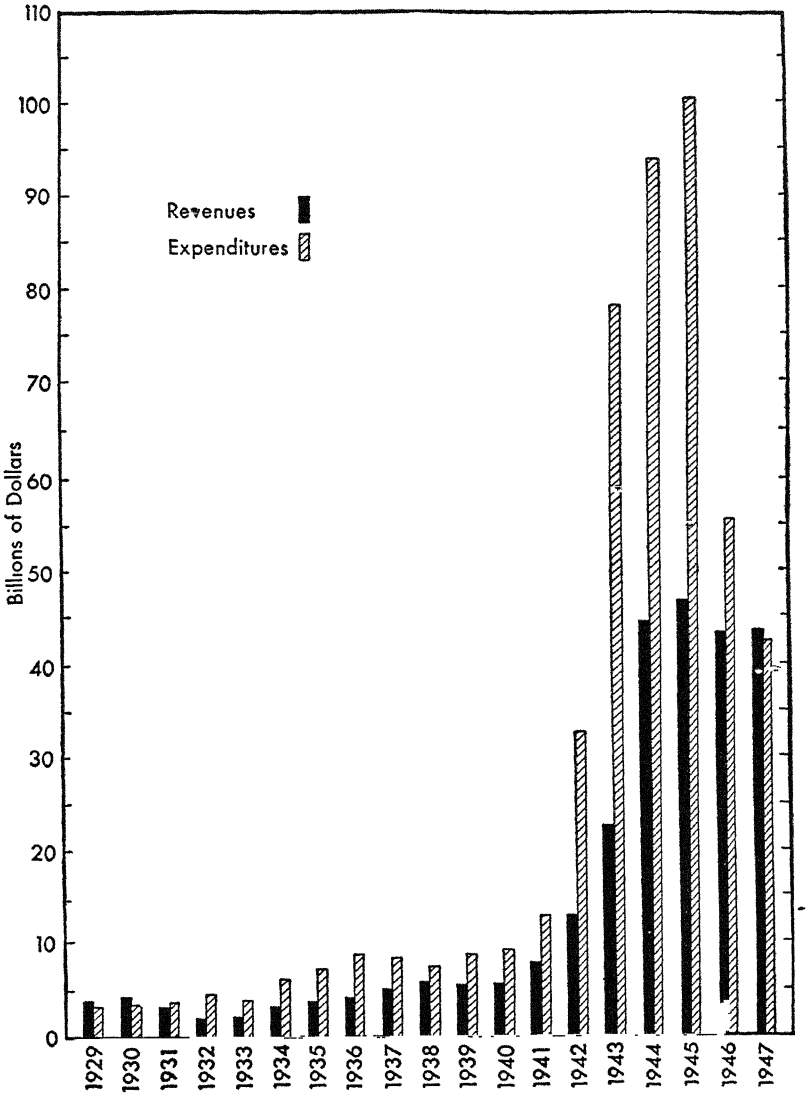


FIG. 34. UNITED STATES GOVERNMENT, RECEIPTS AND EXPENDITURES, 1929–47 [1]

[1] The figures for this chart were taken from the table on page 528.

1931, but 1939 produced a deficit greater than in any year except 1936, and the deficit for 1940 was even higher.

From 1940 to 1946, war dominates the fiscal scene, and deficit financing becomes inevitable. The deficit attained its maximum in 1943 at 56 billions and stayed in that neighborhood through the next two years. The year 1946 shows the effects of the end of hostilities, with a substantial decline in expenditures, a slight falling off in revenues, and a smaller, though still very heavy, deficit of 22 billion dollars.

The budget was balanced for the first time since 1930 when in 1947 there was a surplus of 754 million dollars and in 1948 a surplus of 8,419 millions. However, both revenues and expenditures were in these two years at the unprecedented peace-time level of over 40 billion dollars, and the post-war programs for tax reduction and for expenditures of large sums in aid of foreign countries, for the expansion of the military services, for veterans' benefits, and for other established governmental services provided little encouragement for any reduction in total expenditures or for the continuance of any substantial surplus during the next few years.

CHAPTER XXVII

Interregional Trade and Investment

The bases of interregional trade: General characteristics. All trade is, to a lesser or greater extent, interlocal trade; *i.e.,* trade between persons (physical or legal) domiciled in different localities. When the distance between such localities is significant, we speak of interregional trade.

It is well to keep in mind that all trade is conducted between physical or legal persons, not between geographical divisions or entities. This is true even of state-conducted trade, when the partners are governments of the various countries. The question of motivation appears in each case; whether partners in trade are individuals, corporations, or governments, and whether trade is interlocal, or international.

International trade is that which takes place between parties domiciled in different political states; it represents a special variant of interregional trade, in which the fact that the parties are of several independent governments, each of which has regulatory powers over the economic processes that are currently taking place, creates new and often difficult problems. We shall be concerned, first of all, with the more general case of *interregional* trade; then we shall explore the special problems of *international* trade.

Inhabitants of different regions and localities trade with each other, first of all, in order to acquire goods which they lack in exchange for goods which they can spare. This motive is probably the oldest known to history. It represents even today the hard core of trade which is kept up even under systems of most exacting trade restrictions.[1] But, except when trade is highly restricted through government regulations, this motive is only one of several and accounts for a relatively small part of interregional (and international) commerce.

The second motive for trade is to acquire, from the outside, goods which cannot be produced as well or as cheaply within the given

[1] See Chapters xxix and xxx.

locality or region. A region may have a special advantage over other regions in the production of certain articles. This advantage can be due to the geological conditions of the region, to its climate, to its cultural history, and in the case of national units to its size.

Causes of interregional specialization. To illustrate, let us consider several typical instances:

(1) A region may have coal or iron deposits so much poorer than those of other regions that it is much more economical to import these raw materials from the richer deposits than to extract them from the poorer ones.

(2) A region may have a less fertile soil than another and will find it more advantageous to import wheat, let us say, than to try to grow it.

(3) The climate of a region may not be as favorable to the growth of certain fruit or vegetables as that of other regions; again it may be preferable for this region to import the higher-grade product from elsewhere rather than to grow its own.

(4) A region's past development may not have encouraged the acquisition of certain skills which, in some other region, are more widely present among the population; imports will make available the goods produced by the greater skill of the other region.

(5) The larger size of country A will make it possible for its industry to develop methods of mass production which the industry of country B cannot use owing to that country's smaller size. Country B will then find it advantageous to import the mass-produced commodities from A.

The principle of comparative advantage: Absolute advantage. All of these examples illustrate cases where one region (or country) has an *absolute* advantage over another in the production of some particular good. The uneven distribution of mineral resources, of climates, and of human skills throughout the planet is the most powerful reason for trade, and the bulk of trade is due to the existence of these divergencies. The fact that some areas and countries have an absolute advantage in the production of certain goods while other areas or countries have an absolute advantage in the production of other goods, explains the importance of specialization and of the division of labor. The following hypothetical case will be helpful to a better understanding of this principle.

Let us suppose that country A and country B can produce either wheat or corn, but that A has an advantage in the production of

wheat, while B has an advantage in the production of corn. To make the example concrete, let us say that, with the expenditure of a unit of labor and capital (composed of so many hours of labor, so much machinery, fertilizer, etc.), A can produce either two bushels of wheat or one bushel of corn, while B with the same expenditure of effort (one unit of labor and capital) can produce either one bushel of wheat or two bushels of corn. Then if A and B each spend two units of labor and capital and each produce both corn and wheat, the result will be:

A	2 bu. of wheat + 1 bu. of corn
B	1 bu. of wheat + 2 bu. of corn
Total	3 bu. of wheat + 3 bu. of corn

If each country spends its two units in the production of the good in which it has a real advantage the result will be as follows:

A	4 bu. of wheat
B	4 bu. of corn
Total	4 bu. of wheat + 4 bu. of corn

In this case there is a gain in total product of one bushel of wheat and one bushel of corn, but neither country wants to be without the other grain. Rather than do without corn entirely, A would be willing to spend one unit in producing a bushel. Let us then further assume that A exchanges wheat for corn with B at the rate of one bushel of wheat for one bushel of corn.[1] If two bushels are sold, each country would have wheat and corn, as follows:

A	2 bushels of wheat + 2 bushels of corn
B	2 bushels of wheat + 2 bushels of corn

Comparing this table with the first one, we see that A has gained a bushel of corn and B has gained a bushel of wheat as the result of specializing in production and exchanging the surplus.

Thus the absolute advantage in production leads to specialization and trade, and trade brings benefits to both partners.

Comparative advantage. What would happen if one area or country had an absolute advantage over another in every single item of production? Would there be any possibility for trade between them? Such situations present themselves in real life, and it is a fact of experience that trade develops in spite of them. It is one of the early accomplishments of economic theory to explain how

[1] In these illustrations we ignore the transportation charges.

this can be and to account for trade between two areas, one of which can produce everything more efficiently than the other.

To explain the situation let us assume two other countries, C and D. Country C with one unit of labor and capital can produce either fifteen yards of cloth or one ton of steel. Country D with one unit of labor and capital can produce ten yards of cloth or half a ton of steel. C has an advantage in the production of both cloth and steel, but its advantage in the production of steel is greater. Compared with C, D has no advantage in the production of either commodity, but its "least comparative disadvantage" lies in the production of cloth.

If each country devotes one unit of labor and capital to the production of cloth and two units of labor and capital to the production of steel, the amounts produced by each will be as follows:

C with 3 units	15 yards of cloth + 2 tons of steel
D with 3 units	10 yards of cloth + 1 ton of steel
Total	25 yards of cloth + 3 tons of steel

If however with the same expenditure of effort as before, C specializes in the production of steel, in which its comparative advantage is greater, and D specializes in the production of cloth, in which its comparative disadvantage is less, the amounts will be as follows:

C with 3 units	3 tons of steel
D with 3 units	30 yards of cloth
Total	30 yards of cloth + 3 tons of steel

From the standpoint of total production this specialization has brought about a gain of five yards of cloth, but this is of no particular interest to either country unless the terms of exchange are such that each country can gain thereby. C would be unwilling to sell steel for less than fifteen yards of cloth a ton, since it would be more profitable to manufacture the cloth at home; D, for the same reason, would be unwilling to pay more than twenty yards of cloth for a ton of steel. A rate of eighteen yards of cloth for one ton of steel falls within these limits, and we will assume that this is fixed and that C sells D a ton of steel for eighteen yards of cloth. Then we find, after the exchange has been effected, the following situation.

C	18 yards of cloth + 2 tons of steel
D	12 yards of cloth + 1 ton of steel

Comparison of this with the first table will show that C has gained three yards of cloth while D has gained two yards of cloth. In spite of C's greater productive advantage in both industries, it is worth its while to specialize and exchange products with D. Of course the people of D still have less to show for their efforts than the people of C, but they would have even less if they were unable to specialize on cloth and get their steel by exchange.

The principle of comparative advantage. From these examples may be drawn an important generalization known as the *principle of comparative advantage.* It is to the advantage of a country or region to devote its energies, not to all the lines of production in which it may have superiority, but to those in which its superiority is greatest, provided trade gives it the opportunity to secure the other goods from abroad. Likewise the country which has no line of superiority will find it advantageous to devote its energies to those lines in which its inferiority is least marked, provided the opportunity of trade with other regions is open to it.

This principle of comparative advantage is stated in formal language as follows: *A country gains by concentrating its energies in the production of those goods in which it has the greatest comparative advantage or in which it has the least comparative disadvantage.*

Regional specialization and interregional trade thus promote economy in the use of the productive forces of society. By utilizing the most fertile soils and the richest mines, by turning labor and capital into the industries for which each region is best suited, society as a whole has a greater quantity of consumable goods to share among its members, and each region is enabled to secure a greater quantity of consumable goods than otherwise would be possible.

If however the law of comparative advantage really holds good, why should not each country, or each section of each country, concentrate on the production of one article only and secure all the other goods it needs by exchange? The answer is found in the principle that the existence of a comparative advantage in the production of any commodity depends to some extent upon the quantity produced. There is a limit to the economies of large-scale production, and sooner or later all industry operates under conditions of increasing cost. A rising marginal cost of production is an indication that the comparative advantage in the production of additional units of the good is becoming less.

The United States, for example, has such decided advantages in the production of wheat that it regularly has a surplus of several hundred million bushels for export, but if it were to attempt to supply the entire world with wheat, the cost per bushel would rise far above that which would obtain in other wheat-producing countries, or even in some at least of the countries which now import wheat. Rather than pay the price which in such circumstances would be necessary to remunerate the American farmers for their efforts, foreigners would buy wheat from other sources or grow wheat at home.

The law of diminishing returns therefore sets limits to specialization, and we find few products of any consequence in the world's commerce which are produced in one region exclusively, just as we find few important regions with only one industry.

Applications of the principle of interregional trade: Imports and exports. A proper appraisal of the effects of regional specialization upon productive efficiency makes it possible to appraise correctly the importance, not only of foreign trade in general, but also of imports and exports taken separately. Under the most primitive economic conditions, people exchange goods of which they have an ample supply for goods, obtainable elsewhere, which they lack. The purpose of "exports" here is clearly to pay for the "imports"; to obtain the latter is obviously the intent of the transactions; the former are merely a means of attaining that purpose. This is why it may be asserted that imports are the principal aim of trade, and that we export in order to be able to pay for our imports.

This is a much more sensible and correct view of the significance of foreign (or interregional) trade than is another view, widely represented in the United States and elsewhere, that the most important thing in foreign trade is to export. Popular opinion tends to attach far too much importance to exports, far too little to imports.

This misplaced emphasis goes back to the mercantilists (sixteenth to eighteenth centuries), who regarded an export surplus as being desirable *per se* and considered the influx of precious metals, resulting from an export surplus, as the only real addition to a nation's wealth that is brought about by foreign trade. Basing their argument on somewhat different grounds, modern followers of the mercantilists regard exports as desirable *per se* because they are believed to add to employment, while they hold imports to be

undesirable because, by substituting foreign goods for domestic ones, they are believed to reduce employment.

But as Adam Smith made very clear in his attack on the mercantilists, the fundamental consideration is the relative supply of goods fit for human use. Imports are important because they add to the quantity and variety of goods available to the population of a certain region or country.

As stated before, the country as a whole pays for imports with exports, but this is not the only reason for exports. In a world where the regional division of labor has progressed far, interregional trade is indispensable to reap the benefits of this division of labor. What is true of the division of labor between individuals living in the same community is true of groups of individuals inhabiting various regions of the same planet. By specializing on some one pursuit, the individual produces goods, not only to satisfy his own needs and those of his family, but to satisfy a wider market demand. In the absence of this wider market, he could not specialize so fully.

Similarly the industry of a particular region, under a system of interregional division of labor, will produce to satisfy, not only the needs of that region, but also those of a broader, often world-wide, market. Again in the absence of the wider market, the industry could not reap the benefits of large-scale production. For such industries, exports are as essential as sales at home. For them the foreign markets are a part of the total demand which supports production at greatest efficiency. They do not export in order that others can import; they export in order to dispose profitably of the output which they are capable of producing.

We conclude therefore that imports and exports are both essential to a regional or national economy for two reasons: first, because both follow from the regional division of labor and specialization, leading to more efficient production and higher living standards; second, because the population of a region or country, taken as a whole, pays for imported goods with the proceeds of exports.

Multilateral trade. Primitive barter consisted in a direct exchange of goods for goods between two parties. The goods sold and bought had to be of equal value; otherwise the transaction could not take place. This restricted the scope of trade as well as the scope of the division of labor.

After the invention of money, the operation "sale" became divorced from the operation "purchase." The seller did not have to make immediately a purchase of a value corresponding to that of his sale. As the means of payments acquired a widening acceptability, he did not have to buy from the same person to whom he had made a sale, nor even from a person within the same region or country; the two operations became dissociated in space as well as in time. Important consequences followed.

Country A might buy from country B goods of a value smaller than the value of its exports to B, and at the same time buy from country C goods of a value greater than that of its exports to C. This meant that its trade with B and C, taken separately, would be unbalanced. However the proceeds obtained from the export surplus to B would pay for the import surplus from C, and A's trade with B and C (jointly considered) might balance.

This is what we call "triangular" settlements of accounts, or triangular trade. If we apply the same principle to all the countries of the world, we obtain the notion of *multilateral settlements* and *multilateral trade.*

Here again we have an application to interregional trade of the principles which govern relations between individuals living in the same locality. If we wish to obtain a clear understanding of interregional and international trade, we must always keep in mind that they are but special aspects of trade in general and are governed by the same fundamental principles that govern all other forms of trade. Thus multilateral trade allows the widest possible application of the principle of interregional division of labor and specialization of production; any limitation imposed by government action upon multilateral trade restricts the division of labor and results eventually, in a lowering of standards of living. Yet in the past two decades, the practice of "bilateralism"—*i.e.,* of balancing trade between a country and each of the other countries—has grown in scope and has found many advocates in some areas of the world.

There are two reasons for this revival of primitive barter in the form of bilateralism. One is the breakdown, in the early 1930's, of the system of international payments; the second is the growth of economic nationalism.[1] The establishment and maintenance of an

[1] See Chapter XXIX. Bilateralism offers to the stronger partner the possibility of economic domination of the weaker one, which is important for a nationalistic state engaged in the pursuit of power politics.

effective system of international payments is the most effective way to safeguard, in the future, the survival of multilateral trade.[1]

International prices. One of the important effects of international trade (and generally of interregional trade) is to establish connections on an international plane between prices of goods prevailing in the different local markets. This leads further to the creation of links between the price structures of the various regions and countries, and between their respective economic developments.

International or world prices are prices for specific commodities, which prevail at a given time all over the world. These prices are not uniform, owing to transportation costs.[2] Moving goods from one point to another on the earth's surface involves costs, and these costs have to be paid by the buyers of the transported goods. Other things being the same, wheat will be more expensive in England than it is in the wheat belt of the United States, or in Alberta, Canada. Coal will be more expensive in Italy than in the Saar, and so forth. When the costs of transportation are very high in comparison with the price of the commodity transported (both per unit of the commodity), we cannot, strictly speaking, refer to a world price of that commodity; when such costs are relatively small, we can.

There is a further limitation to the notion of international prices. It is only meaningful and useful when applied to commodities of great importance to the world economy and of standard quality. It is the existence of standard quality which makes possible the equalization of price (allowing for the differentials due to transportation costs). When in a given country the price of a standardized good, say wheat, goes up above its world level plus transportation charges, profits can be made by shipping wheat to that country, and these shipments will bring prices down to their world level. Alternately, if the price of wheat is below the world level in some other country, it will be advantageous to export wheat from that country, which again will bring about a price equalization.

This process of simultaneously buying in the cheapest market and selling in the highest is called *arbitrage*. It is only applicable to standardized merchandise and it represents the process by which world prices are established.

[1] See Chapters XXVIII and XXIX.

[2] We shall disregard here artificial barriers to international trade, such as tariffs, quotas, etc., since the effect of these barriers is to interfere with the establishment of world prices.

The ability to specialize to advantage is evidenced by a relatively low cost of production, which enables the producers of one region to undersell those of other regions, and the opportunities for the extension of regional specialization and trade are brought to light by differences in the prices of the commodity in the various markets. Under competitive conditions price serves as an automatic control over regional division of labor and interregional trade.

Industrialization and interregional trade. One of the fairly widespread misconceptions about international trade consists in viewing that trade as essentially an exchange of raw materials produced in certain areas of the world for manufactured goods produced in other areas. According to that view it is unwise for industrialized countries to encourage or assist the industrial development of countries whose products are principally raw materials, because by so doing they will deprive themselves of future markets for the output of their own industries. It is necessary to make clear this point of view because, if accepted and acted upon, it would have a stifling effect upon the division of labor and upon standards of living.

Facts point in the opposite direction. The volume of foreign trade grows as the economic development of a country takes forward strides. In 1938 out of a total value of United States exports of 3,094 million dollars, 521 million went to Great Britain, 575 million to the industrial countries of Continental Europe, and 476 million to Canada. Compare the 146 million dollars' worth of exports to the rest of Europe (not counting the Soviet Union) and 70 million to the Soviet Union. It stands to reason that an industrialized India [1] could absorb more American goods than the 38 million dollars' worth imported in 1938, and an industrialized China [2] more than the 52 million dollars' worth imported.

The fallacy we are dealing with here arose presumably from the fact that the British trade in the first half of the nineteenth century consisted very largely of exporting manufactured goods and importing raw materials, but this was a very exceptional situation. All through history the exchange of manufactured articles for other manufactured articles, and of raw materials for other raw materials, has represented a significant part of interregional and international commerce.

The industrialization of underdeveloped areas of the globe re-

[1] Including Burma and Ceylon.
[2] Including Manchuria.

sults in demand for imported machinery and equipment, for imported technical skills, and for imported manufactured consumer goods. Eventually the growing countries begin to export manufactured goods, in the production of which they have either absolute or comparative advantages. The industrial growth of the backward areas is an incentive to the further growth of the advanced areas.

In each case growth involves changes and readjustments in the economic structures of the respective countries. It is a matter of experience that the advanced countries are often resistant to change, and that trade barriers are erected to avoid change. These are reactionary practices, an aspect of economic nationalism, and they are eventually paid for in terms of lower standards of living—in advanced and backward areas alike—than would prevail were the process of growth and change allowed to proceed undisturbed.[1]

The structure of world trade.[2] Some statistical data will be helpful in appraising the importance and the characteristics of international trade. There is no way of compiling comparable data on

WORLD IMPORTS AND EXPORTS, 1938

GROUP	IMPORTS		EXPORTS	
	In millions of dollars	Per cent	In millions of dollars	Per cent
British Commonwealth	8,040	32.7	5,857	26.7
United States and overseas territories	2,366	9.6	3,228	14.7
France and overseas territories	1,781	7.2	1,332	6.1
Netherlands and overseas territories	1,262	5.1	1,130	5.2
Belgium, Belgium Congo, and Ruanda-Urundi	802	3.3	776	3.5
Italy and overseas territories	752	3.1	562	2.6
Spain and overseas territories	169	0.7	119	0.5
Portugal and overseas territories	153	0.6	82	0.4
Union of Soviet Socialist Republics	268	1.1	257	1.2
Germany	2,222	9.0	2,162	9.8
Japan, Korea, Formosa	728	3.0	817	3.7
All other countries	6,040	24.6	5,597	25.6
World total	24,583	100.0	21,917	100.0

[1] See League of Nations, *Industrialization and Foreign Trade.* New York: Columbia University Press, 1945; Brown, A. J., *Industrialization and Trade.* London: Royal Institute of International Affairs, 1943; Staley, Eugene, *World Economic Development.* Montreal: International Labor Office, 1944.

[2] Cf. League of Nations, *The Network of World Trade.* New York: The Columbia University Press, 1942. Figures on world trade used in this chapter, if not otherwise identified, are quoted from this important volume.

interregional trade taking place within a given country, as the statistics are rarely if ever available. The table on page 541 shows how some of the important political subdivisions of the world shared in 1938 in world imports and exports.[1]

The following table gives an idea of the composition of world exports by groups of merchandise in 1937:

	Per cent of total exports
Foodstuffs and live animals	22.1
Materials, raw or partly manufactured	39.0
Manufactured articles	38.9
Total	100.0

It will be seen that nearly two-fifths of world trade consists of manufactured goods; the figure would be higher if we included semimanufactured goods as well. Since tariffs and other barriers to trade tend to limit trade in manufactured articles more strictly than trade in raw materials and foodstuffs, the proportion of the former in the total trade of the world would be substantially higher than shown by the 1937 figures if the world were to adopt free trade.

Foreign trade of the United States:[2] **Volume.** It is not possible to give here detailed information about the foreign commerce of the United States, but the broad data contained in the tables that follow will provide the reader with the most important facts pertaining to the quantity, general character, and development of United States imports and exports. In view of the exceptional disturbances of the war years and of the unsettled situation at the end of the war, the statistics given deal with the prewar period.

During the war years the exports of the United States grew tremendously, reaching the value of 8 billion dollars in 1942, 12.8 billions in 1943, and 14.1 billions in 1944. Imports remained considerably more stable; the corresponding figures for 1942, 1943, and 1944 are: 2.8 billion dollars, 3.4 billions, and 3.9 billions.

[1] It will be noted that the figures for world imports and world exports do not coincide in our table. This is due to the inclusion in imports of transport costs between the frontiers (or ports) of the importing and exporting countries, amounting in 1938 to about 12 per cent of the value of world exports. (*The Network of World Trade, op. cit.*, p. 16.)

[2] Cf. Hal B. Long and Associates, *The United States in the World Economy*, U. S. Department of Commerce, Economic Series No. 23. Washington: Government Printing Office, 1943.

Value figures give, of course, only an approximate idea of the evolution of foreign trade on account of the fluctuations in prices.

FOREIGN COMMERCE OF THE UNITED STATES

(In millions of dollars)

	EXPORTS	IMPORTS	EXCESS OF EXPORTS (+) OR IMPORTS (−)
1800	70.9	91.0	− 20.1
1820	69.7	74.4	− 4.7
1840	123.7	98.2	+ 25.5
1860	333.6	353.6	20.0
1880	835.6	667.9	+ 167.7
1900	1,394.5	849.9	+ 863.0
1910	1,744.9	1,556.9	+ 188.0
1920	8,228.0	5,278.5	+ 2,949.5
1925	4,910.0	4,227.0	+ 683.0
1928	5,129.0	4,091.0	+ 1,037.0
1930	3,843.0	3,061.0	+ 782.0
1932	1,611.0	1,323.0	+ 288.0
1934	2,133.0	1,655.0	+ 478 0
1937	3,345.0	3,084.0	+ 261.0
1939	3,123.0	2,276.0	+ 843.0

During the great depression, for example, prices fell so heavily that the value of imports and exports was more considerably affected by this factor than by the actual decline in the volume of goods traded.

Commodity character of foreign trade. The accompanying table

COMMODITY CHARACTER OF EXPORTS AND IMPORTS IN PERCENTAGES OF TOTAL EXPORTS AND IMPORTS OF THE UNITED STATES

COMMODITY	IMPORTS		DOMESTIC EXPORTS	
	1860[a] Per cent	1937[b] Per cent	1860[c] Per cent	1937[d] Per cent
Crude materials	11	32	68	22
Crude foodstuffs	13	14	4	3
Foodstuffs wholly or partly manufactured	17	15	12	5
Manufactures for further use in manufacturing	10	21	4	21
Manufactures ready for consumption	49	18	11	49
Total	100	100	100	100
Total (in millions)	$354	$3,012	$316	$3,295

[a] General imports for year ending June 30.
[b] Imports for consumption for calendar year.
[c] Year ending June 30.
[d] Calendar year.

gives for selected years the commodity character of the foreign trade of the United States.

These figures tell the story of a revolutionary change in the character of the foreign trade of the United States, indicating the growth in relative importance of the nation's manufacturing industries. In 1860 the nation exported 72 per cent of its goods in crude form. In 1937 all but 25 per cent of the domestic exports were partly or wholly manufactured, and the percentage of consumable manufactures had increased from 11 to 49. On the import side the converse is seen to hold true. The percentage of imported manufactures ready for consumption declined from 49 in 1860 to 18 in 1937; the percentage of goods partly fabricated but requiring further processes increased; and the percentage of raw materials imported increased greatly.

The direction of foreign trade. The following table shows the destination of exports and the sources of imports for great trading areas in recent years:

FOREIGN TRADE BY GREAT TRADE REGIONS

(Annual averages. Values in millions of dollars)

	NORTH AMERICA				SOUTH AMERICA		EUROPE		ASIA		OCEANIA		AFRICA	
	Northern		Southern											
Year	Value	Per cent	Value	Per cent	Value	Per cent	Value	Per cent	Value	Per cent	Value	Per cent	Value	Per cent
Exports:														
1910–14	320	14.8	181	8.4	121	5.6	1,350	62.3	121	5.6	48	2.2	25	1.1
1921–25	627	14.3	445	10.1	297	6.8	2,318	52.7	499	11.3	141	3.2	70	1.6
1926–30	820	17.4	403	8.4	448	9.4	2,236	46.8	574	12.0	177	3.7	110	2.3
1932	246	15.3	119	7.4	97	6.0	784	48.7	292	18.2	37	2.3	36	2.2
1934	308	14.4	178	8.3	162	7.6	950	44.5	401	18.8	57	2.7	77	3.6
1936	392	15.9	225	9.2	204	8.3	1,043	42.5	399	16.2	79	3.2	114	4.7
Imports:														
1910–14	119	7.0	229	13.5	207	12.2	836	49.5	259	15.3	17	1.0	23	1.3
1921–25	397	11.5	514	14.9	421	12.2	1,049	30.4	943	27.3	54	1.6	71	2.1
1926–30	480	11.9	460	11.4	546	13.5	1,211	30.0	1,193	29 6	53	1.3	91	2.3
1932	181	13.7	157	11.9	201	15.2	389	29.4	362	27.4	8	0.6	24	1.8
1934	238	14.4	161	9.7	229	13.8	489	29.6	490	29.6	15	0.9	33	2.0
1936	381	15.7	237	9.8	292	12.0	718	29.6	708	29.2	36	1.5	51	2.1

The quantities are given both in terms of value and in percentages of the totals of exports and imports respectively, thus portray-

ing both the absolute and the relative importance of the different regions of the world in American foreign trade.

We cannot attempt an exhaustive study of the trade relationships revealed by this table, but attention should be called first of all to the great, though declining, importance of Europe in American foreign trade. In 1936 she took about 43 per cent of the total exports of the United States and provided nearly 30 per cent of the imports. Of the total exports to Europe, 40 per cent was in the form of crude materials; foodstuffs amounted to 15 per cent; semi-manufactured goods, to 16 per cent; and finished manufactures, to 29 per cent. As might be expected, imports from Europe consisted chiefly of manufactured goods, both finished and partially finished.

The importance of the trade with North America, attributable almost entirely to the trade with Canada, is striking. Commerce with the southern part of North America—in which the Cuban trade bulks largest—shows a decline, both absolutely and relatively, after a period of slight growth. The trade with the countries of South America, on the other hand, has grown materially since 1910. In part this may be ascribed to the privileged position which the United States enjoyed during two World Wars, but in the main it is a reflection of the complementary nature of the production of the United States and of many countries of South America. A very natural basis exists for the exchange of the coffee, nitrates, tin, hides, etc., of South America for the manufactured goods of the United States.

Another feature deserving of mention is the position of Asia. In the twenty-seven years covered in the table, American imports from Asia have more than doubled, and exports to Asia have almost tripled. Raw silk, rubber, tin, furs, hides and skins, jute, and other raw materials predominate in the imports from Asiatic countries, while exports to Asia consist mainly of manufactured goods.

Imports. The table on the next page shows the value of the ten leading articles imported into the United States in 1936.

Judged on the basis of value, these ten items are the most important articles imported. A mere glance at the list will suggest how essential most of these products are to the economic life of the people. But it would require a painstaking study of the hundreds of different items coming from all parts of the world to enable one to appreciate fully the extent to which economic life is dependent on foreign sources of supply. In addition to the goods already

LEADING IMPORTS OF THE UNITED STATES IN 1936

(In millions of dollars)

Crude rubber	158.7
Cane sugar	157.9
Coffee	134.0
Paper and paper manufactures	110.1
Raw silk	102.4
Paper base stocks	98.9
Vegetable oils	85.3
Furs and fur manufactures	81.6
Chemicals and related products	80.0
Tin	75.5
Total	1,084.4
Total imports	2,423.0

mentioned, we should find many minerals and other raw materials —such as nickel, manganese, antimony, jute, and flax—which in America are produced not at all, or in insufficient quantities. There would be many manufactured goods which perhaps the American people could produce, but which they can get with less effort by trade, such as fine woollens from England and watches and lace from Switzerland. We should discover also many articles of food—tea, coffee, cocoa, and out-of-season fruits and vegetables—which could be secured at home only at prohibitive cost.

Of recent years there has been much talk of the advantages of a self-contained nationalistic economy, to be attained presumably by the erection of prohibitive tariff barriers. A study of the problems raised by the protective tariff will be taken up in a succeeding chapter, but it may not be out of place here to point out that a study of its imports soon convinces one that even the United States, with its wide expanse of territory and its diverse natural resources, could not hope to proceed very far in that direction unless the people were willing to alter substantially their mode of life and submit to a distinctly lower standard of living.

Exports. The table on the opposite page gives the values of the ten leading exports from the United States in 1936.

The sum total of the exports is a part of the price which has to be paid for the imports and other services from abroad. The exports are the goods in the production of which American advantages are greatest. By devoting time and energy to their production the imported goods are secured on easier terms than would be the case if they were to be produced at home (assuming this to be possible),

or if capital and labor were to be devoted to making other goods in America for sale abroad.

LEADING EXPORTS OF THE UNITED STATES IN 1936

(In millions of dollars)

Cotton, raw	361.0
Machinery	334.9
Automobiles	240.2
Petroleum and its products	179.0
Tobacco, unmanufactured	137.3
Chemicals and related products	116.9
Iron and steel-mill products	111.9
Fruit and nuts	80.6
Coal and coke	56.6
Copper, ore, and manufactures	50.5
Total	1,668.9
Total exports	2,456.0

It is because a country's exports represent the most efficient means of acquiring certain desired goods that every country is properly concerned with the condition of its export trade. Any factor, domestic or foreign, which interferes with the free outflow of exports decreases the purchasing power of a nation in foreign markets and tends to lower the standard of living.

There is another reason why a nation may view the volume and character of its export trade with some anxiety. While it is true

UNITED STATES EXPORTS OF LEADING COMMODITIES IN RELATION TO PRODUCTION

	Percentage exported			
	1914	1929	1933	1936
Tobacco, leaf[a]	47.2	41.2	39.3	33.4
Cotton[b]	62.6	54.8	65.6	56.1
Wheat[c]	19.7	17.9	5.4	2.5
Copper, refined	54.8	36.2	41.0	32.1
Refined mineral oils	37.9	19.6	11.9	9.9[e]
Industrial machinery	—	13.3	11.2	13.3
Agricultural implements and machinery	—	25.1	29.8[d]	10.6
Automobiles	4.5	12.0	7.3	8.0
Office appliances	23.6	30.2	26.1	23.1

[a] Based on production figures for the preceding calendar year and exports for the year ended June 30.

[b] Based on production figures and exports for the year ended July 31.

[c] Exports include flour.

[d] Data are for 1931.

[e] Data are for 1935.

that production for the foreign market is on the average a fairly small part of total production, in particular industries the foreign-market production may amount to a very large fraction of total output. The table on the preceding page gives for selected industries the percentage of total production in the United States which was exported in four selected years.

These are the more conspicuous examples, but many industries could be cited which have customarily exported up to 10 per cent or so of their total output. Quite naturally anything which interferes with the foreign market of such industries is a serious matter, for the loss of 10 per cent, or even less, of the total market may mean the difference between producing at a profit and at a loss. The loss of foreign markets is of immediate practical concern, not alone to the particular producers immediately involved, but to the nation as a whole. If foreign markets are closed to an appreciable extent, the depressing influence is not confined to the export industries but tends to be generalized and cumulative. In the course of time labor and capital, thrown out of use in the export industries, can be diverted to other uses and equilibrium achieved anew, but not without irreparable loss.

World trade after Second World War. The reconstruction of world trade was a much more difficult task at the end of the Second World War than it was at the end of the First World War. By 1939 the international division of labor, never fully restored after 1918, had become seriously disorganized. The breakdown of international trade in the depression of the early 1930's was followed by the growth of bilateralism, first adopted by the German government and then by several others, as an aspect of the economic mobilization for war. The war of 1939–45 itself completely disrupted the remaining channels of international commerce.

The concentration of foreign trade in the hands of government departments made much headway during the war, and today the great issue is between free-enterprise foreign trade and state-controlled trade. The former leans inevitably towards the multilateral pattern; the latter, towards bilateralism.

We shall deal in a later chapter with the question of economic nationalism; here it is important to indicate that the movement towards state trading tends strongly to increase the scope of restrictions which national governments place on the operations of international trade. The United States has constituted itself a lead-

ing champion of nondiscriminating, multilateral trade, conducted by private individuals and firms. Leaving aside the political aspects of the problem, its economic aim is to take trade out of power politics and make it an instrument for the creation of wealth and prosperity rather than of military predominance.

The wider the area over which trade follows the lines of economic advantage and not those of political expediency, the greater will be the benefits derived by the peoples of the various countries. As far as the United States is concerned, this accomplishment is hampered by popular acceptance of the fallacious idea that exports are more important to well-being than imports. The American tariff still interferes with importing, especially of many manufactured articles produced abroad which the public might otherwise enjoy.

Foreign investments: Their nature. We speak of foreign investments when capitalists or public bodies of one country invest money in the securities or businesses of another country. As in the case of domestic investments, we must distinguish between credit and ownership investments. The former take the form of bonds or other evidences of debt; the latter are operating investments or "direct" investments in business enterprise.

Loans may be made by private persons or governments either to private foreign borrowers or to foreign government, including provinces or states and municipalities as well as national governments. Direct or ownership investments are, however, generally private. Each of these types of financial transactions presents its own unique problems but at the same time is affected by the principles common to all international credit transactions.

Purposes of foreign borrowing. An international investment transaction may be initiated in either the capital-importing country or the capital-exporting country. Private persons or governments of one country may seek capital from another country because of the absolute or relative scarcity of capital in the domestic market. A private enterprise borrows or sells stock abroad to obtain funds which are either not available at home or are available at home only on onerous terms. Governments likewise borrow abroad because this is the only, or the cheapest, source from which to obtain capital.

While private companies seek capital abroad to finance their operations, governments borrow for a variety of purposes. Under unusual circumstances, they may borrow to reinforce their mone-

tary systems against adverse conditions in the balance of their foreign trade, or to provide for reconstruction following a major war. Normally their borrowings are undertaken to finance their economic development or their military preparedness. Since our present interest is in the normal functioning of international finance, the following discussions will be directed primarily to investment transactions undertaken to promote the long-term economic development of the borrowing country.

Economically the obtaining of capital abroad enables a country to import more than it exports for a more or less extended period. Current exports of the less developed areas of the world are insufficient to provide the funds to pay for the imports required for their development, and foreign capital makes possible for them the excess of imports over exports which their desired development requires. It therefore follows that a country receiving foreign capital will concurrently receive a net inflow of goods.[1]

A state-socialistic nation with a powerful and ruthless government, if its own rate of capital formation is too slow to provide the desired degree of economic development, may so regulate its domestic economy as to speed up that rate rather than borrow abroad. By enforcing long hours of work and by keeping down the standard of living of the people, it may increase both the total national product and the portion thereof devoted to capital formation. Thus it exacts from its own people the means to provide roads, dams, power plants, industrial equipment, and capital in general.

Such a government may also obtain the imports needed for its economic growth, without resorting to foreign loans, by the forced sale of exports and by restricting drastically so-called nonessential imports, through the use of stringent controls on imports and by dumping or subsidizing of exports. Of course these alternatives to normal interchange of capital through international investments are available only to a strongly entrenched authoritarian government and can be exercised only at the sacrifice of individual economic and political freedom and of a possible higher standard of living.

The need for foreign investment funds is therefore greatest under

[1] Although a country may at the same time both import and export capital, it is the difference between the values of the two movements which determines for this discussion whether the country is receiving or providing capital.

relatively free international trade. Then it is that free people endeavor through foreign loans to improve their standards of living by advancing industrialization. This possibility of improved living conditions is today more strongly realized than ever before, and most governments are pledged to policies aimed at its realization.

Purposes of foreign lending. In the making of foreign investments, the private investor and the government are motivated by essentially different purposes. The private investor constantly seeks opportunities for those uses of his accumulated savings which offer the highest possible returns consistent with the degree of security and extent of control which he considers desirable. He buys foreign investments when their terms are to him more attractive in these respects than alternative investment opportunities at home.

Governments however are influenced in lending abroad by other considerations. These considerations, though they can involve economic purposes and objectives, are more apt to be of a political nature.

During the nineteenth century, not only were the prospective returns of foreign investments more attractive and their risks less forbidding, but private capital itself was more venturesome, so that much of the industrialization and economic growth of that century was financed with private funds. During the past three decades, foreign investments have declined in popularity, and private capital has become more timid, as the spread of nationalism and socialism over the world has limited the opportunities for private venture capital. Governments and their lending agencies have turned toward this function, and their lending has thus tended to become more economic in character.

When once a country has, with the aid of foreign governmental loans, made a substantial beginning in its industrialization, investment opportunities in continuing the process become more attractive to private investors. As between loans and direct investments, the experience of the United States during the past twenty years would indicate that the direct investments offer the greater promise.

The transfer of payments. The soundness of domestic investments depends upon the solvency of the borrower or of the company in which funds have been invested on an ownership basis. In the case of foreign investments, an additional consideration—and a very important one—relates to the mechanism of foreign payments.

Within any given country there is no problem of transferring payments from one region to another. This fact is due to the existence of a common monetary unit and of complete freedom of trade.

It is in transactions between separate countries that the problem of transferring payments arises. The national currency of a country usually has buying power only within the confines of that country; movements of precious metals provide only a temporary means of settling accounts; and eventually payments between nations must be carried out by means of an export surplus achieved by the paying country and of an import surplus achieved by the country receiving the payment.

Under conditions of multilateral trade, there is no need for the debtor country to send to the creditor country goods of the value of the payment due, nor need the creditor country accept goods directly from the debtor country. It may well be that the respective economic structures of the two countries are such that the one cannot absorb an adequate amount of goods produced in the other. It is sufficient for the success of the transfer that the debtor country achieve an appropriate export surplus in its trade with the rest of the world and that the creditor country have a corresponding import surplus in that international trade. The foreign exchange markets of the world will then take care of the rest.

The international investment position of the United States. The depression of the 1930's and the Second World War have played havoc with the world economy in general and with foreign investments in particular. England, formerly a large creditor, has come out of the war as a debtor nation on balance. The creditor position of the United States has been greatly weakened. Indeed by the end of 1944 this country was creditor on balance on long-term account only, but was a debtor on balance if we include short-term liabilities. This was, of course, an entirely temporary situation, due on the one hand to losses on foreign investments, and on the other to the inflow of short-term funds from abroad—funds sent here for safety just before the war and after its outbreak. Since the end of the war, the United States Government, both directly and through the Export-Import Bank, has extended substantial foreign loans, and public and private foreign investments have increased.

The United States is today the only major creditor country of the world, and the only one likely to expand its creditor position considerably in the next decades. The other countries in this category

include Canada, Switzerland, and Sweden. Although several other countries, while remaining debtors on balance, are likely to resume foreign investments once their own reconstruction is completed, the position of the United States as the world's leading source of investment capital for export is unique.

The International Bank for Reconstruction and Development. To effect the transition from government loans to private investments, the International Bank for Reconstruction and Development was established in 1945. Its purposes may be stated as follows:

"(i) To assist in the reconstruction and development of territories of members by facilitating the investment of capital for productive purposes, including the restoration of economies destroyed or disrupted by war, the reconversion of productive facilities to peacetime needs and the encouragement of the development of productive facilities and resources in less developed countries.

"(ii) To promote private foreign investment by means of guarantees or participations in loans and other investments made by private investors; and when private capital is not available on reasonable terms, to supplement private investment by providing, on suitable conditions, finance for productive purposes out of its own capital, funds raised by it, and its other resources.

"(iii) To promote the long-range balanced growth of international trade and the maintenance of equilibrium in balances of payments by encouraging international investments for the development of productive resources of members, thereby assisting in raising productivity, the standard of living, and conditions of labor in their territories.

"(iv) To arrange the loans made or guaranteed by it in relation to international loans through other channels so that the more useful and urgent projects, large and small alike, will be dealt with first.

"(v) To conduct its operations with due regard to the effect of international investment on business conditions in the territories of members and, in the immediate postwar years, to assist in bringing about a smooth transition from a wartime to a peacetime economy."[1]

The bank established in the pursuance of these objectives is an experiment in coöperative international finance. Its authorized capital was set at 10 billion dollars, to be subscribed by the governments of member states. The subscription of the United States is 3,175 million dollars, that of Great Britain 1,300 million.

[1] Cf. U. S. Department of State, *United Nations Monetary and Financial Conference. Final Act and Related Documents,* Publication 2187, Washington, 1944; Heilperin, Michael A., *International Monetary Reconstruction,* New York, American Enterprise Association, 1945; National Planning Association, *The Stakes of Bretton Woods,* Washington, 1945.

Only 20 per cent of the subscribed capital is paid in; the rest constitutes a guarantee fund to be drawn upon in case of need. The bank can guarantee or extend loans to the limit of its subscribed capital; these loans to be made, except in special circumstances, "for the purpose of specific projects of reconstruction and development"—an expression which includes loans for the purpose of monetary stabilization.

The bank can either guarantee loans issued in the private capital market, or give loans from its own funds or from funds borrowed by itself in the market. It is possible that the operations of the bank may do more than supplement private investments and replace purely governmental loans. Assuming that its credit is established and it has made a reputation for itself, the bank may, through its guarantees and participations, reopen the doors to the international flow of private capital. As a multi-currency institution, operating in the currencies of all its members, it would naturally be interested in stable international monetary relations.

CHAPTER XXVIII

Foreign Exchange and the International Balance of Payments

Initial assumptions. In our approach to the study of foreign exchange it will be convenient to eliminate for the time being certain complicating factors of recent years. The difficulties which confront the student of contemporary foreign exchange arise partly from the general abandonment of the gold standard and partly because of a variety of controls, exercised by governments or by central banks in pursuance of governmental policy, affecting exchange operations or transactions which would involve the exchanges.

We shall examine some of these problems in the next chapter, but for the present we will assume (1) that the countries which we are considering are on the gold standard, implying thereby that all forms of money within each country are ultimately redeemable in the country's standard gold coins and that the rights of free coinage, export, and import of gold may be freely exercised by private persons, and (2) that there are no formal exchange controls which interfere with the operations of a competitive market in foreign exchange. Such were the conditions which existed with respect to the exchanges of all the leading industrial nations in the period prior to the First World War, and it is within the bounds of possibility that they may again be attained at some time in the future.

Financing exports. Our first problem—one which is much the same whatever the monetary systems on which the exchanges are based—is to investigate the mechanism which is available for financing exports and imports of goods. The simplest approach to this is by way of the study of typical examples of export and import transactions. Let us therefore assume that Richards & Green of New York has sold Harris & Company, Ltd., of London, a certain number of bales of cotton for £1,000. Now it is evident that the ordinary forms of currency used in domestic trade are not available for payment in this case. The universal currency for international

payments is the bill of exchange, and we may assume, what would almost always be true in such a case, that the terms of the sale have called for a bill of exchange to be drawn by Richards & Green on Harris & Company, Ltd. The bill will be in the following form:

No. 726 New York, December 29, 1929
Pay at sight to the order of the New York Bank One Thousand Pounds Sterling.....
(signed) Richards & Green
To Harris & Company, Ltd. London, England

The bill of exchange is an instrument of such importance that we must pause here to formulate its precise definition, which is as follows: *A bill of exchange is an unconditional order in writing, signed by one person (called the drawer) directing a second person (called the drawee) to pay a third person (called the payee) or his order a certain sum of money at some definite time in the future.* It will be noted that the example given above conforms with this definition.

When Richards & Green ships the cotton it secures from the steamship company, or possibly from a railroad company, if the shipment is to be made from an inland warehouse, a bill of lading, which not only shows that the company has received a certain number of bales of cotton and has agreed to transport them to England, but also serves as the evidence of ownership in the cotton. Anyone in possession of a properly endorsed bill of lading can secure the cotton from the steamship company, and it can be secured in no other way. Marine insurance must be taken out on the shipment in order that the owner may be protected if the cotton suffers loss or damage during the sea voyage.

Richards & Green next takes the bill of exchange, with these two documents (bill of lading and insurance policy) and possibly others attached, to its bank, the New York Bank, for sale. The bank knows nothing about the reliability of Harris & Company, Ltd., but it knows that Richards & Green is an old, well-established house with a reputation for financial straightforwardness. If it buys the bill of

exchange it will acquire a credit instrument bearing the signature of a responsible officer of this firm and having the same standing as a promissory note signed by the same officer. If Harris & Company, Ltd., fails to meet the bill when presented for payment, the bank can enforce payment from Richards & Green.

Furthermore the bank acquires possession of the bill of lading, which will give it the title to the cotton. Now cotton is a staple article, not perishable, and commanding at once a definite price in all the important markets of the world. If the worst should happen and the bank could not collect from Richards & Green, it could sell the cotton and still not lose much. Relying therefore on the responsibility of Richards & Green and on the cotton as collateral security, it purchases the bill of exchange at the rate prevailing in the market at the moment for bills of that character. The rate, let us say, happens to be \$4.84[1] per pound sterling, and Richards & Green therefore receives \$4,840 in cash. The exporting firm has thus received payment for the cotton in American dollars immediately after shipment and has no further interest in the transaction, save for its contingent liability.

The bank now sends the bill, with the documents still attached, to the London Bank, with which it has correspondent relations, with instructions that it be presented at once to Harris & Company, Ltd., for payment and that the proceeds be credited to the account of the New York Bank. Harris & Company, Ltd., thereupon pays the bill in British currency and receives the bill of lading, which entitles it to claim the cotton from the agents of the steamship company.

We may summarize these stages in the financing of the export of cotton as follows: (1) Richards & Green draws a bill of exchange on Harris & Company, Ltd., for £1,000. (2) Richards & Green sells the bill of exchange (with documents attached) to the New York Bank for \$4,840. (3) The New York Bank sends the bill and the documents to its correspondent in London, the London Bank. (4) The London Bank collects £1,000 from Harris & Company, Ltd., and surrenders the bill of lading. (5) The New York Bank's deposit account with the London Bank is credited by £1,000.

Financing imports. A similar example will aid us to visualize the several steps involved in the financing of an import. We assume

[1] The student will realize that this rate reflects conditions of 1929 when both Great Britain and the United States were on the gold standard.

that White & Company of New York has purchased £800 worth of cotton goods from R. J. Littlefield of London. It would be quite in accord with commercial practice for Littlefield to draw a bill of exchange on White & Company, thus reversing the process described in the previous section, but trade practice frequently places the responsibility of making payment, as well as of securing payment, on the American merchant.

We may assume then that in this instance White & Company is required to place Littlefield in possession of £800 in cash or its equivalent in London. This it can do by purchasing from its banker a bill of exchange such as the following, often called a demand or sight draft.

No. 918 New York, January 5, 1930

At sight pay to the order of R. J. Littlefield
.....Eight Hundred Pounds Sterling.....
and charge the same to our account.

(signed)
New York Bank

To
The London Bank

The New York Bank sells this bill of exchange to White & Company at the prevailing rate for bankers' demand drafts, which we will assume to be $4.87. The draft therefore costs White & Company $3,896. White & Company sends the draft by mail to Littlefield, who presents it at the London Bank and receives £800 in cash, which amount is debited to the account of the New York Bank.

Other international transactions. While the financing of the import or export of goods usually forms the largest part of international financial transactions, there are innumerable other transactions between individuals, corporations, and governments in one country and individuals, corporations, and governments in other countries which require the transfer of funds from one country to another and thus give rise to a demand for, or a supply of, bills of exchange.

As an example we may assume that a New York investment house has been authorized to sell an issue of bonds for the British government. In all probability the investment house would use the proceeds from the sale of the bonds to purchase from a New York banker a bill of exchange on a London banker drawn to the order of the British Treasury. It would thus add to the demand in New York for bills of exchange on London. When the interest on these bonds became due the British Treasury might authorize its agent in New York to draw a draft on the Treasury for the amount of the interest payments. To secure dollars for actual payment to the bondholders, the agent would offer this draft for sale to a New York banker and thus add to the supply of bills of exchange in New York on London.

It is unnecessary for us to examine each of the many international transactions—shipping, banking, insurance, personal services, and the like—performed for or by persons in a given country, but it is essential to realize that we can know little about the total demand and supply of bills of exchange until we have a complete picture of all such transactions.

The role of the banks in exchange. The part which the bankers play is that of middlemen. It would indeed be possible for the importers to buy bills of exchange from exporters and to remit these to their creditors. In the past this has been done, but today virtually all bills are, in the first instance, either sold to or bought from the bankers. The banker has a wide circle of customers importing and exporting every known ware and dealing with all the markets of the world, and he is able to buy and sell drafts in all currencies and in all amounts at a moment's notice. An importer of goods from China can secure a draft payable in Shanghai or Hongkong with no difficulty, and furthermore he knows and the Chinese exporter knows that the risk of loss inherent in a banker's draft is much smaller than in a draft drawn by one commercial house on another. Because of these advantages it is not surprising to find the banker at present the focal point of the foreign exchange market.

Middlemen always expect to sell what they buy and to make their profit by charging a higher price for the good than they paid for it. Bankers do not differ from other middlemen in this respect. Their willingness to buy bills of exchange and thereby to acquire bank credits abroad arises from the fact that they expect to be able

to utilize these credits by selling their own drafts at a higher rate. It is the margin between the banks' buying rate and their selling rate which gives them their profit in exchange transactions.

The exchange market. The exchange rate responds in so sensitive a manner to the play of the forces of demand and supply, and there is such a vast quantity of transactions in foreign exchange in New York and other centers, that it is frequently assumed that there must be an organized market similar to the stock and produce exchanges where the rate of exchange is determined. This is an error; there is in New York no formal organized exchange for dealing in foreign exchange. The bulk of the business is concentrated in the hands of a relatively few large banks and exchange houses. Each bank fixes the rates at which it will buy and sell exchange on all the financial centers with which it has dealings, but each house keeps so closely in touch with the activities of the other houses that there is seldom any material discrepancy between the rates charged by the various dealers on the same class of bills.

In foreign exchange dealings between two countries on the gold standard there is no official rate of exchange. Each bank posts the rates at which it is willing to buy and sell bills of exchange, but it frequently departs from its posted rates both on bills of large amounts and unquestioned character and on bills of dubious standing.

The rate of exchange: General principles. Thus far we have been concerned chiefly with the technical facilities which modern business has at its disposal for the transmission of funds from one country to another, but we are equally interested in the forces which determine the *rate of exchange.* This may be defined as the *price paid in the home country for one unit of the money of the foreign country payable in that country.* For example the rate of exchange in New York on London is the amount of American money that must be paid in New York for a bill of exchange calling for the payment of one pound sterling in London.[1]

The first step in the study of the rate of exchange is to determine what is known as the *mint par of exchange.* This is a simple ratio

[1] Rates of exchange may equally well be expressed the other way around, as the amount of the foreign money payable in the foreign country which may be obtained in the home country for one unit of its money. Thus before the First World War the rate of exchange on France was often stated as so many francs per dollar. This is not the usual method, however.

derived by dividing the fine gold content of the standard unit of one country by the fine gold content of the unit of another. Italy and Rumania, for example, used to have gold coins containing the same number of grains of fine gold; the mint par of exchange between them was therefore one; *i.e.,* one lira was equivalent to one leu. Formerly the sovereign of Great Britain contained 113.00116⅔ grains of fine gold, and the gold dollar of the United States contained 23.22 grains of fine gold. The ratio was 113.00116⅔:23.22 or 4.8665+. Hence the British sovereign contained 4.8665+ times as many grains of fine gold as the dollar, and the mint par of exchange was $4.8665+ per pound sterling. This means that if 10,000 gold sovereigns of full weight were to have been taken to the United States mint, they would have contained sufficient fine gold to have furnished the gold content of $48,665+.

The rate of exchange, however, is not the same thing as the mint par of exchange. Occasionally the rate of exchange quoted in the market stands at the mint par, and when it does so, we say that exchange is "at par." But this rarely occurs, and then only by chance. The mint par of exchange furthermore does not determine the rate of exchange; its significance lies, as will appear, in the fact that it helps to set limits to the fluctuations of the actual rates.

The rate of exchange is a price and, in common with competitive prices in general, is determined by demand and supply. The demand for bills of exchange comes primarily from the importers and others who are under the necessity of making payment in foreign countries. The supply comes primarily from the exporters and others who have money available to them in foreign countries. The bankers are the middlemen who, buying and selling, bring together the forces of demand and supply. In accordance with the general laws of price, the greater the demand for bills (other things being equal), the higher the rate of exchange; the greater the supply (other things being equal), the lower the rate. In terms of international trade, when imports are large compared with exports, the rate of exchange tends to be high; when exports are relatively large, the rate tends to be low.

The gold points. But the rate of exchange differs from most competitive prices in one important respect; between countries on the gold standard it can fluctuate only within definite limits. Having risen to the upper limit, the rate will stand there, no matter how great the demand, until changed conditions send it down.

Having fallen to the lower limit, it can go no lower and will eventually rise.

To make this clear, let us assume that, because of the transactions of a certain short period, businessmen in America owe £50,000,000 for imports received and are owed only £25,000,000 for exports. By buying the bills of the exporters, American bankers can acquire credits in London of £25,000,000 and be in position to sell their own bills to this amount to the importers. But this will leave £25,000,000 of the debts of the American importers unpaid. As this discrepancy between the volume of bills which importers and others want and the volume offered to the bankers for sale begins to be manifest, bankers will tend to charge a higher price for their own bills of exchange in order to protect their foreign balances, and individual bankers, each of whom will want to get as large a share of the possible credits abroad as he can, will bid against each other for the bills of exchange which are offered for sale by exporters. The rate of exchange will be forced up gradually, and the higher rate will tend to discourage some who want to make payments and perhaps encourage others to anticipate the drawing of bills of exchange.

But if the basic forces at work are strong enough, the bankers will soon find their foreign balances exhausted, with importers still trying to buy bills of exchange. In this situation the bankers in America, if they are not to cease operations entirely, are forced to resort to other expedients to build up their foreign bank balances. They might acquire balances against which they could draw by borrowing from a bank in England or by exporting securities for which there is an international market, but, granted our assumption of a free gold market, they would in all probability export gold from the United States for this purpose.

At a mint par of exchange of $4.866 per pound sterling, gold equivalent to £1,000 can be purchased in the United States for $4,866. It will obviously cost something to cover the shipment of gold—to take care of crating, freight, insurance, and the like. Assuming that these expenses amount to 2½ cents per pound sterling, then the total cost of placing £1,000 in London would be $4,891. The bankers could therefore offer to sell bills of exchange drawn on banks in London at $4.891 per pound, and the importers who were under obligations to make immediate payment would have to pay this rate. True, they could ship the gold themselves,

but the banks can do this much more cheaply, and competition among different banks forces them to do so.

This rate we call the *gold exporting point*, determined by the cost in New York of transferring one gold sovereign to London. It is the highest price which will be paid in one country for one unit of a foreign currency payable in the foreign country, and no matter how great the demand, bankers will continue to furnish bills of exchange to all comers at that rate so long as gold is available for shipment.

In like manner we may determine the *gold importing point.* Let us suppose that the supply of sterling bills of exchange—that is, bills drawn on London in pounds sterling—increases greatly, without any corresponding change in the demand. As we have seen, a supply of bills of exchange arises when exporters and others having debts to collect draw bills of exchange against the foreign debtors and attempt to sell them to bankers, and a demand arises when importers and others having debts to pay to foreigners go to the bankers to buy bankers' drafts.

In this case, therefore, bankers are likely to have large credit balances abroad for which there is no call, and if the situation continues, they will refuse to buy bills except at rates which will permit them to transfer their foreign balances to replenish their dwindling domestic cash resources. The cost of shipping gold from England to America is slightly higher than in the other case, for in this instance the bankers lose interest during the time of shipment; whereas when gold is shipped from New York, the bankers can sell drafts to go by the same steamer and so avoid tying up their funds.

If we assume the cost of shipment to be 3 cents per pound sterling, bankers will be willing to buy bills of exchange at the rate of $4.836. With a draft of £1,000 a banker can obtain 1,000 gold sovereigns, bring them to the United States, and sell them for $4,866. The cost of doing so being $30, he will recover $4,836, the amount he paid for the bill. This rate, $4.836 per pound sterling, is known as the gold importing point, and at this rate bankers will be willing to purchase an unlimited quantity of bills of exchange, no matter how great the supply.

The gold points then are determined with respect to the mint par of exchange and fix the limits within which the rate of exchange may fluctuate. If the rate rises to the upper gold point, gold will be exported; if it falls to the lower gold point, gold will be

imported. The rise or fall in the rate is determined by the relation of demand and supply to each other, which in turn rests on the relation between our obligations to foreigners and the obligations of foreigners to us. If these were just equal, the rate would presumably be at or very near the mint par of exchange, and there would be no tendency for gold to flow at all; if at any time one greatly outbalances the other, gold will flow out to create the necessary foreign bank balances, or gold will flow in as bankers find this the only way to utilize their foreign balances.

Indirect exchange. It is scarcely necessary to remind the reader that the demand for and supply of bills of exchange, which determine exchange rates and gold movements, relate to the total of all international transactions of the country concerned and not to the account between just two countries. If the latter were the case, there would be almost continual shipping in and out of large quantities of gold, since it is scarcely conceivable that the debits and credits of any two countries should exactly offset each other except as a rare coincidence. As a matter of fact, when the majority of the countries of the world were on the gold standard, international trade was always conducted on a clearing house basis, in accordance with which all the obligations of the residents of one country to residents of all other countries were offset against all claims against foreigners. There was no international clearing house, but the result was accomplished with remarkable certainty and smoothness through the foreign exchange markets of the world.

In this system London and sterling exchange played an important role. The leading position of Great Britain in world trade, the solidity and flexibility of her monetary and banking system, and the supremacy of London in international short-term and long-term lending as well as in other financial services gave to London a dominating position in international finance, which only recently she has lost. These circumstances gave to sterling exchange the character of an international currency. For many years bills on London were in all markets a most acceptable form of exchange.

An illustration of the workings of indirect exchange, in the days when sterling exchange served as an international currency, can be seen in the trade between Great Britain, Brazil, and the United States. The American people normally imported more goods and services from Brazil than they sold to her, they exported more to Great Britain than they purchased from her, and the people

of Great Britain sold more to the Brazilians than they purchased from them.

If each country had been forced to settle with each separately and without regard to its financial obligations to other countries, the United States would have been required to send a large sum in gold annually to Brazil, Brazil in turn would have had to settle in gold with Great Britain, and a third shipment of gold would have been necessary from Great Britain to the United States. Or if gold could not have been shipped, Americans would have cut down their imports from Brazil until they balanced the exports to her, Brazil in turn would have decreased her imports from Great Britain until they were in balance with her exports to Great Britain, and American exports to Great Britain would likewise have been decreased. The second alternative would have meant obviously a tremendous decrease in international trade, with unfortunate reactions on international division of labor.

The obligations created by this three-cornered trade were, however, settled much more simply. America's excess of exports to Great Britain gave bankers in the United States balances in London banks which were utilized in drawing bills of exchange in favor of Brazilian exporters. In due course these drafts were sold by the Brazilian exporters to bankers in Brazil and sent to London for collection. Brazilian bankers therefore acquired balances in London banks against which drafts were sold to Brazilian importers of goods from Great Britain for transmission to the British exporters.

It is possible, though very unlikely, that the surplus balances acquired from American exports to Great Britain might have been exactly the amount necessary to pay for the excess of imports from Brazil; it is more likely that Americans would still have had a balance in London, which could have been utilized to pay for an excess of imports from Colombia, let us say.

Limits to the flow of gold: General. As we have seen, gold flows out of a country when the volume of international payments it has to make to other countries is so much in excess of the volume other countries have to make to it that the rate of exchange is driven to the gold export point; but unless there are fundamental causes of disequilibrium at work gold will not continue to flow out indefinitely. For several reasons the fluctuations in the exchange rates which bring about an inflow or an outflow of gold tend to be self-correcting.

Effects of gold movements on price levels. In the first place a substantial flow of gold tends to affect the price levels in the gold importing and exporting countries in such a way as to alter the profitableness of importing and exporting goods. How this serves to bring the flow of gold to an end can best be illustrated by a simple example.

Let us consider two countries, A and B, dealing exclusively with each other, and let us suppose that A in a given year imports $100,000,000 in gold because of an extraordinarily large favorable balance of payments. The gold comes into the hands of bankers in the first place and either is gradually put into circulation or serves as a reserve for other forms of currency. Sooner or later its influence is likely to be felt in credit expansion, increased business activity, and a rise in the general price level. Similarly the withdrawal of $100,000,000 from circulation in B may cause, other things remaining the same, a restriction of credit, a slackening in business activity, and a fall in the general level of prices.

While gold was flowing into A the rate of exchange must have been down to the gold importing point. Now however there comes a transformation in the exchange market. The rise in the price level in A will make A a better market for the sale of goods from B but a poorer market from which to buy. Similarly B will become a cheaper market from which to purchase and a poorer market in which to sell. Imports from B into A will therefore tend to increase, while exports from A to B will tend to decline. The increase in imports means an increase in the demand for bills of exchange, while a decrease in the supply is caused by the falling off of exports. The simultaneous increase in demand and decrease in supply will have the effect of raising the rate of exchange to some point above the gold importing point in A, thus putting a stop to the further inflow of gold.

Likewise an outflow of gold from country A to country B would be only temporary, for the export of gold would reduce the price level in the gold exporting country and raise the price level of the gold importing country. In the first country the export of goods and services would be stimulated and the imports checked until the increase in the supply and the decrease in the demand for bills of exchange would cause the rate to fall below the gold exporting point, so stopping the outflow.[1]

[1] The automatic operation of this check can of course be nullified if the country receiving the gold "sterilizes" it; *i.e.*, prevents it from entering, directly or

Gold movements and credit transfers. Monetary contraction and expansion in response to gold movements has been obtained in practice through changes in interest rates in the gold-losing and gold-receiving country. Interest rates in gold-receiving countries would fall, thus tending to expand bank credit; rates in gold-losing countries would rise, with the opposite effect on the quantity of money in the form of bank credit.

But these differential changes in interest rates reacted not only upon monetary circulation, but also and very directly upon international transfers of bank balances. If because of a flow of gold from the United States to England, the rate of interest is increased in New York and reduced in London, some London bankers will wish to acquire balances in New York in order to earn the higher rate of interest. Funds may be transferred from London to New York, either by the purchase of dollar exchange in London or by authorizing a New York banker to sell sterling exchange on London.

Thus there would result an increase in the demand for dollar exchange in London, or an increase in the supply of sterling exchange in New York, or both. In any case the sterling rate in New York would tend to drop, and if there were a substantial transfer of short-term capital from London to New York, it would drop below the New York gold export point. In the operation of the international gold standard such international movements of short-term funds were of great importance in limiting exchange fluctuations and the flow of gold—perhaps of greater importance than price and trade adjustments.

Reaction of exchange rates upon imports and exports. Aside from the regulating force of the international flow of gold and the transfer of short-term funds, the exchange rates themselves react upon imports and exports and tend to bring about an equilibrium. This relation will become clear if, placing ourselves in the position of the merchant whose business brings him into the exchange market as a buyer or seller of bills, we consider the effect of the exchange rates upon his transactions.

Let us assume that this typical merchant is exporting to England and that sterling bills are selling at a high premium in New York.

indirectly, into circulation and consequently from affecting the price level, and if the country losing the gold neutralizes the effect of this on its own price level by a more liberal extension of bank credit.

With exchange in this position, the merchant may make an extra profit from his sales in England. His normal trading profit is computed in the sale price of his goods and is therefore included in the face of the sterling bills which he draws against the English buyers. In turning his bills into dollars at the prevailing high rate of exchange, a second profit is derived from the fact that each pound sterling is worth more dollars than normally. Now if this merchant is in active competition with other exporters, as is the normal case, the profitable market for American goods in England will result in lower sale prices to the benefit of English buyers, and this situation will increase exportation, because of the principle that more can be sold at a low than at a high price.

Obviously the position of the American importer of English goods is the reverse of this when sterling exchange sells at a premium in New York. Since payment of goods imported from England is made by the purchase of sterling bills in New York for remittance to the English exporter, the American importer finds his profit cut down by the premium on exchange. The premium becomes an additional cost in the business of importing English goods, and the prices of these goods will rise and their import diminish.

These illustrations should suffice to make it clear that the relation of the rates of exchange to the items entering into international exchange is one both of effect and of cause. The rise and fall of the rates is controlled fundamentally by the demand and supply of bills of exchange. But when an excess of imports forces foreign exchange to a premium in any market, the high rate tends to restrict imports and encourage exports and thus works toward a restoration of equilibrium in the foreign trade. Conversely when a large volume of exports makes the rates on foreign bills fall to a discount, imports are stimulated and exports restricted, which again tends to bring the trade into balance.

With an international gold standard in operation, the rates, confined as they are within the bounds of the gold points, can rise or fall so little that their influence upon the balance of payments may be imperceptible. It must be borne in mind, however, that this narrow range exists only because gold is permitted to flow in sufficient volume to settle international accounts; when gold is prevented from supplying this corrective, there is no definite limit to the fluctuations of the rates, and a premium or discount in the

rate of exchange may therefore have greater influence on international transactions.

We conclude that gold cannot flow in one direction, either into or out of a country, in large amounts for an indefinite period, because the flow of gold provides its own automatic checks through its effect on the price level, the exports and imports both of goods and of short-term capital, and the rate of exchange. Temporary discrepancies in the balance of payments are settled from time to time in gold, but economic forces work in the direction of reducing the flow of gold to the minimum, and while in a given year a country may have considerable net imports or net exports of gold, consideration of a longer period will show the imports and exports of gold tending approximately to offset each other. The tendency over a period of time is for the debits and credits—the amounts owed by residents of a country to foreigners and the amounts owed by foreigners to them—to equal each other. We pay foreigners for goods and services by the export of our own goods and the rendering of services to foreigners; in the last analysis the actual transfer of money becomes trifling.

The balance of payments: Its composition. In order to secure a clear picture of the international financial transactions of a given country over a period of time, a sort of international balance sheet is frequently drawn up. In the credit column would be found, for the period under consideration, exports of mechandise as well as all other transactions, such as the sale of securities abroad, services performed for foreigners, and the like, which could give rise to a supply of bills of exchange in the country in question and hence give its bankers credits in foreign banks. In the debit column we would include imports of merchandise and all other transactions, such as purchases of foreign securities, tourist expenditures abroad, services performed by foreigners, and the like, which could give rise to a demand for bills of exchange in this country and so permit its bankers to utilize its credits in foreign banks. Since it is the effect on gold movements that we are now studying, we do not include among the credit and debit items the export or import of gold itself.

If these total credits exceed the total debits, it means that the volume of bills offered to bankers exceeds the volume demanded of them, the rate of exchange will be at the gold import point, and gold will be imported to settle the balance. This situation is known as a *favorable balance of payments*. On the other hand, if total debits

exceed total credits, conditions of demand and supply will push the rate of exchange to the gold export point, and gold will flow out of the country. This is known as an *unfavorable balance of payments.*[1]

The international balance sheet of the United States. The picture of the international financial relations of a great nation can be made clearer by a glance at the table on the following page, derived from figures compiled by the United States Department of Commerce, showing the international balance sheet of the United States for the year 1928.[2]

In this table every transaction which creates in the United States a claim against some foreigner is considered a credit. Such items furnish the supply of bills of exchange on foreign countries, or a demand in foreign countries for bills on the United States. Thus the item "Earnings of long-term investments" means that individuals or corporations in the United States acquired credits abroad amounting to $817,000,000 because of interest and dividends from their holdings of foreign securities.

Similarly every transaction which gives rise to a claim of some foreigner against an American is placed in the debit column of our account. Thus shipping services performed by foreigners cost $227,000,000, and the cashing of the drafts to which the payment of this amount gave rise decreased the foreign bank balances of Americans by this amount. Such items furnish the demand for foreign bills of exchange, or the supply of bills on the United States available to foreigners.

Fallacies of the "favorable balance of trade." The early mercantilists believed that if the merchandise exports of a country exceeded in value its merchandise imports, precious metals would flow in to settle the balance. Since to them the precious metals were especially desirable forms of wealth, this situation came to be called a "favorable balance of trade." Similarly when the value of merchandise imports exceeded the value of the exports, the condition was known as an "unfavorable balance of trade," since it was ex-

[1] The terms "active" and "passive" balance of payments are frequently used nowadays in place of the terms "favorable" and "unfavorable," which tradition has handed down.

[2] The balance sheet for 1928, rather than one of a later year, is considered more useful for purposes of illustration since it antedates the months in 1929 when speculation was at its height, the years of depression, and the abandonment of the gold standard.

ESTIMATED BALANCE OF INTERNATIONAL PAYMENTS OF THE UNITED STATES, 1928[1]

(In millions of dollars)

Commodity Trade	Credits	Debits
Merchandise exports and imports	5,129	4,091
Silver	87	68
Ship chandlering, sales of bunker coal and oil, etc.	98	62
Unrecorded parcel-post shipments	20	20
Adjustments in above items	—	256
Total commodity trade (as adjusted)	5,334	4,497
Miscellaneous invisible items		
Freight payments and receipts	143	227
Tourist expenditures	168	782
Ocean-borne passenger traffic	89	—
Earnings of long-term private investments	817	252
Earnings of short-term interest and commissions	65	107
Immigrant remittances	28	217
War-debt receipts of United States Treasury	210	—
Other United States government receipts, payments, foreign representations here	53	110
Missionary and charitable contributions, etc.	—	52
Motion-picture royalties	70	6
Insurance transactions	80	70
Miscellaneous	46	53
Total miscellaneous items	1,769	1,876
Movement of private long-term capital		
New American investments abroad (net)	—	2,070
Reduction of previous American investments abroad	853	—
New foreign investments in the United States	1,704	—
Reduction of previous foreign investments in the United States	—	1,223
Total private, long-term capital items	2,557	3,293
Movement of short-term capital		
Net change in international banking accounts	—	226
Pure cash items		
Gold shipments	561	169
Changes in ear-marked gold[a]	68	188
Total	629	357
Grand total, all items[b]	10,289	10,249

[a] Gold held in the vaults of a bank, but segregated and marked in such a way as to show it is the property of another bank.

[b] It is apparent that the credits and debits do not exactly balance, but it should be realized that many of the figures represent estimates only, and necessarily so, since there is no law compelling tourists to inform the government of the sums they spend abroad or compelling American shippers to certify as to the amounts they pay foreigners for ocean freight. It must also be remembered that there is no settlement date upon which a country's account for international payments must be balanced. The balance of payments is a continuing account, and even if our statistics were exact to the last dollar we should not expect an exact balance.

[1] Source: U. S. Bureau of Foreign and Domestic Commerce. *Estimated Balance of International Payments of the United States, 1928.* Washington: Government Printing Office, 1929.

pected that precious metals would have to be exported. Even today popular belief clings firmly to these notions, and it seems as logical to many of our contemporaries as it did to the mercantilists to stimulate exports and repress imports in order that prosperity may be brought to a country through the import of the precious metals.

There are three fallacies involved in this concept of a favorable balance of trade. The first lies in conceiving the exports and imports of merchandise to be the only items in the balance sheet; the second involves the concept of gold as the most desirable of all imports; and the third is the supposition that a steady inflow of gold can be induced by artificial means.

The first of these fallacies has now been disposed of in connection with our development of the nature of the international balance of payments. A few words on the second fallacy will be in order here.

Is gold a peculiarly desirable import? Many persons would immediately answer this question in the affirmative. They experience a feeling of satisfaction when the statistics show that gold is coming in, they "view with alarm" an outflow of gold, and they would favor governmental action to encourage the import of gold.

In the first place, it should be evident that gold obtained through international trade adds nothing to the total wealth of the nation, since it is obtained only by the sacrifice of other goods presumably of equal value. If a country lets these other things go in order to get gold, it presumably gains in utility through the import of gold. But it just as truly gains when it exports gold in order to obtain other things which it wants more.

As regards the use of gold as money, the important fact is that, within limits, the absolute quantity of money is pretty much a matter of indifference. Through the adjustment of prices, a given volume of exchanges may be consummated with either a small or a large quantity of money; what occurs is that the price level is low in the one case, high in the other.

It must be recognized however that there are times when the movement of gold is a matter of more than usual concern to the treasury and bank officials of a nation. A country which is with difficulty maintaining the gold standard quite naturally looks upon the export of gold in any considerable amount as a matter of the utmost consequence and as justification, so it thinks, for any restrictive measures which may so operate on the international balance of

payments as to prevent it. Likewise a country which is contemplating a return to the gold standard in one form or another finds the acquisition of gold reserves essential and will undoubtedly consider gold so valuable an import as to warrant substantial sacrifices in goods and services.

In general however there is nothing, except for the needs of a monetary system at critical periods, to set gold apart as a peculiarly desirable import and nothing to be gained by forcing its import in excess of what would enter under the normal operation of international trade. Under conditions such as existed prior to the First World War, the flow of gold is an automatic thing, controlled by economic forces which result in a state of equilibrium which is for each nation the most favorable situation possible in view of all its other circumstances.

Ignorance of this principle lay at the foundation of the third of the aforementioned mercantilist fallacies. This particular fallacy has been to a considerable extent exposed in our development of the automatic operation of the natural forces which normally control the international flow of gold. The futility of attempts to create a favorable balance of payments by means of tariffs and other artificial restrictions of imports will be shown in Chapter xxx.

EXERCISES

1. Mr. Jones, an exporter of grain from the United States to France, sells a shipment worth $10,000 to M. Guerne, agreeing to receive payment by drawing on the buyer. Draw the bill of exchange, and trace the steps by which the transaction is handled.
2. Suppose that a New York City bank will pay $4.80 per pound sterling for a demand draft drawn by an American exporter on an English importer.
 (a) What price would the bank pay for a time draft payable 90 days after sight, if the rate of discount were 4 per cent? 2 per cent?
 (b) How does a decline in the rate of discount affect the "spread" between the prices of these two types of bills of exchange?
3. Suppose that a New York City bank asks $3,600 for a bankers' sight draft in the amount of 10,000 rupees drawn on Calcutta. How much would the buyer have to pay for a 10,000 rupee draft payable 90 days from date? Assume that the rate of interest is 4 per cent, and the mail time from New York to Calcutta is 30 days.
4. Construct an international balance sheet from the following items, and determine the balance of trade and the balance of payments.

Exports of merchandise	$2,400,000,000
Shipping services rendered by foreigners	80,000,000
New American investments abroad	100,000,000
Interest received by American investors	65,000,000
Imports of merchandise	2,100,000,000
American tourist expenditures abroad	320,000,000
New foreign investments here	600,000,000
War debt receipts	5,000,000
Immigrant remittances	35,000,000
Foreign tourist expenditures here	80,000,000
Interest paid to foreign investors	75,000,000
Gold	——?

5. Explain how each of the following tends to affect the price of foreign exchange in this country:
 (a) Decreased export of American cotton
 (b) Development of an American merchant marine
 (c) Stoppage of war debt payments to this country
 (d) Decline of immigrant remittances
 (e) Increased imports of Canadian lumber
 (f) Stoppage of interest payments due American investors
 (g) Redemption of American securities held abroad
 (h) Loss of confidence in the stability of a foreign currency, causing foreigners to increase their investments here
 (i) Increased sales of American automobiles abroad
 (j) Decline of American tourist expenditures abroad.

CHAPTER XXIX

Contemporary International Monetary Problems

Breakdown of the international gold standard: Prewar equilibrium. In the chapter immediately preceding, we have followed the workings of foreign exchange as it prevailed in the period before the First World War, when the important nations were on a gold standard. It was a period not only of stable exchange relations but of substantial equilibrium as well. Changes of course were taking place, and there were periods of excessive business activity and periods of recession, but there was a mutual adjustment of fundamental factors in national and international life which we do not have today. Price levels were in equilibrium. International trade followed lines determined by a fairly well-settled international division of labor and was not so constricted by universal high tariffs as today nor so distorted by frequent changes in tariffs. International investment proceeded along conservative lines under the leadership of the experienced investment houses of London.

This international equilibrium was brought about largely by the normal operations of the gold standard, which provided, as we have seen, automatic checks to factors which might make for disequilibrium. Abnormal factors operating through the price level or the balance of payments could be taken care of by exchange operations without causing disaster either to the exchanges or to the internal economy of a country.

The operation of these automatic controls tended to keep the rates of exchange normally within the gold points, but it should be noted also that central banks often intervened in the market to bring the rate of exchange down from the gold exporting point when it was believed that this rate was the result of temporary dislocations in demand and supply. The standard method was for the central bank to raise its rate of rediscount. A rise in the rate of rediscount at the Bank of England, for example, would make it more profitable to leave funds in London than to transfer them to New York, or it might even induce funds to come there, thus alter-

ing the demand and supply of bills of exchange on that centre in such a way as to check the outflow of gold or to reduce its volume.

Postwar exchange conditions. It is unnecessary to emphasize the internal and external dislocations brought about by the First World War, but it is important—even though we can do no more than mention a few of the major ones—to show why these dislocations and others, the product of the postwar period, have given the world its heritage of disordered exchanges.

In the first place, it is clear that the practical problem of financing war needs made both the abandonment of the gold standard and some degree of inflation almost imperative. Inflation came about as a result of unbalanced budgets and was resorted to in degrees which varied with the financial strength of the government concerned and its ability to raise revenue either by taxation or by bond issues. It meant of course the printing of inconvertible money and the expansion of deposit currency through government borrowing from central banks for the purpose of meeting current governmental expenditures.

The United States did not find it necessary to balance its budget by the direct emission of paper money, but it did not escape inflation, which was occasioned by the importation of millions of dollars in gold and by increase in the issue of federal reserve notes (a comparatively new form of money at that time) and in the volume of bank credit made possible by this influx of gold.

Two facts of special significance emerge from this monetary disorder. The first is that exchange rates became erratic and subject to violent fluctuations, leading to a vast increase in the risk entailed in foreign transactions and to an unaccustomed volume of speculative operations in the exchanges. The second is that inflation proceeded to different degrees in the inflating countries, so that the natural relation of price levels to each other was distorted.

Preoccupation with war and the production of war materials led both to the growth of huge war industries in the European countries and also to the decay of other industries, notably some which had been most successful in the export trade. Lack of shipping space, as well as lack of goods for export, transferred much of the export trade of European countries to the United States. The creation of new national boundaries by the peace settlement separated regions which had been economically unified, and this dislocation was intensified by an increase of tariff barriers which a new and

fiery nationalism demanded. International trade was thrust from its accustomed channels and impeded in ways which brought disastrous results on the balance of payments of various countries.

Finally the heritage of war debts between the Allied and Associated Powers and the enormous reparations payments demanded by the Allies added another strain to international finance, a strain intensified and indeed made intolerable by the unwillingness of many countries to receive payments in goods and services.

Irredeemable money and the par of exchange: What is a "natural" rate of exchange? Before discussing the problems of the exchanges in the postwar period it is essential to understand the relation existing between the money of a country on an inconvertible paper standard and that of a country on a gold standard, and also the relation between the respective units of two countries both of which are on an inconvertible paper standard. When the British government in 1919 "unpegged" sterling exchange in New York—that is, when it ceased to give it an artificial price by a continuing offer to buy sterling at a fixed rate—it was said that sterling was free to find its "natural level." What is a "natural" rate and why is another rate an "unnatural" one?

The purchasing power par of exchange. The most satisfactory answer to this question is found in what is called the *purchasing power par of exchange,* which may be defined as *a rate of exchange between two currencies which expresses an equivalence in the purchasing powers of the respective units in the home countries.* Thus if $3.00 were to be assumed to be the purchasing power par of exchange between the dollar and the pound when both the United States and Great Britain were on a depreciated paper standard, it would mean that the pound had three times as great a command over commodities in general in Great Britain as one dollar had in the United States.

The purchasing power par of exchange is of course not a fixed par but fluctuates with changes in the purchasing powers of the respective units, as measured by changes in the index numbers of the price levels. If the price level in the United States were to rise while the British price level remained unchangd, the purchasing power par would move upward from our assumed figure of $3.00, while if the British price level were to rise and the American price level to remain stationary, the purchasing power par would decline from $3.00.

Purchasing power par and market rate of exchange in a period of stability. The purchasing power par of exchange is a useful concept in any period of economic equilibrium and when we are considering two countries whose trade with each other is sufficiently large to influence the general level of prices in each country. Market rates of exchange will actually tend to revolve around this par of exchange. A rate higher than our assumed par of $3.00 would mean that sterling exchange was at a premium in New York. Imports into the United States would consequently be checked and exports would be encouraged. An increase in the supply of sterling exchange and a decrease in the demand in New York would cause the rate of exchange to fall towards the purchasing power par of exchange. Similarly a discount on sterling in New York would affect the currents of commerce in such a way as to bring the rate of exchange up towards the purchasing power par.

Purchasing power par and market rates in a period of inflation. In a period of rapid economic change, however, the adjustment of prices to disturbing factors is very slow, both nationally and internationally, and in such a time the market rates of exchange may bear little direct relationship to the purchasing power par. Let us suppose for example that the United States should inflate its currency. The immediate effect would be a rise in the rate of foreign exchange; it would take more American dollars to buy a pound sterling. Eventually there would also result a corresponding rise in the prices of American goods. If this effect were to come instantaneously, the monetary inflation would have no influence on export trade. A pound sterling would cost more in American dollars, but American goods would be worth more in dollars, so that in terms of domestic goods foreign exchange would be no more costly than before, and there would be no tendency to check imports. From the foreigner's side, American dollars could be obtained more cheaply in terms of, let us say, British pounds, but American goods would be correspondingly higher priced in dollars, so that there would be no added inducement to buy American goods; American exports would therefore not be encouraged.

Actually it takes considerable time for the general price level to adjust to monetary inflation. In the meantime, therefore, the high exchange rate, caused by monetary inflation, exerts its normal influence to check imports and stimulate exports. The effect of this on the rate of exchange, other things being equal, would be to check

the speed with which it was rising. Nonetheless, so long as inflation proceeded, we would expect the rate of exchange to be continuously at a premium over purchasing power par of exchange. When inflation ceased and the domestic price level had adjusted itself to the new quantity of currency in circulation, we should expect the rates of exchange again to be determined somewhere around the purchasing power par.

Limits to the usefulness of the purchasing power par concept. The concept of the purchasing power par of exchange is subject to some theoretical weaknesses. For example, exchange rates result from the demand and supply situation in the market, and this is affected not only by payments resulting from trade but to a marked degree by financial—credit and capital—transactions whose volume is not controlled by price levels. Even with respect to trade itself, the value of exports and imports does not depend entirely upon changes in the index numbers which record price changes for all commodities or for selected groups of commodities; income factors and other considerations play a part.

While such factors would reduce to some extent the usefulness of the purchasing power par if it could be determined accurately, a more important consideration is that the practical determination of the purchasing power par of exchange is beset with apparently insuperable difficulties; at least thus far no method of computing it has been discovered which is accepted as theoretically sound. The experimental method is the only one available, and it is the one which must be used by officials in charge of exchange stabilization funds.

On the basis of indexes of price levels and of all other pertinent economic data, some rate must be chosen as a tentative par of exchange. The officials must then by appropriate operations in the exchange market strive to keep the market rates of exchange from diverging too widely from this assumed par of exchange. If it is found that as a result of these operations the market rates are continuously discouraging exports, it is clear that the tentative or assumed par is below the desired neutral par of exchange and the officials must revise their tentative par of exchange. By a series of shifts of this character they may hit upon the proper par, which will not be very distant from the purchasing power par, but even so it should be noted that their difficulties are not at an end, for the par is or may be a moving par.

The return to the gold standard: In general. The reader is already aware of the many practical and psychological advantages to be obtained from the automatic controls of the gold standard, augmented as might be desired by other methods of control. At the close of the First World War these advantages appeared so compelling that nearly every country looked forward to the time when it could feel it was in position to resume the redemption of its paper money in gold and permit the free export and import of gold.

Many countries succeeded in this during the ten years following the war. We can do no more than touch briefly on the history of stabilization in Great Britain, whose experiences in returning to gold are instructive in view of the fact that this problem may confront many countries in the future.

Stabilization of the pound. Partly for reasons of national pride and partly because of pressure from the class of investors in government securities, Great Britain returned in 1925 to the gold standard at the old gold content of the pound. Thereafter the gold or mint par of exchange of the pound to the dollar was much higher than the natural level of the two currencies, as determined by the purchasing power parity. And the consequent overvaluation of the pound in terms of the dollar introduced serious elements of instability into the British economy.

At this time also British public finance was in difficulties. Heavy increases in government expenditures, with inadequate taxation, were causing deficits which contributed to loss of confidence in the financial stability of the British government. These influences, together with other disturbing factors, undermined and finally overthrew the gold standard.

To understand the influences of the monetary factor, we have only to consider the effect of the new level of rates on exports and imports. Assume that a fair degree of stability had been achieved in the internal and external relations of Great Britain prior to revaluation, and then ask what effect the new exchange rates would have on exports and imports. Clearly they would raise the prices of all British export goods in foreign markets and decrease their sale; quite as clearly the price in Britain of foreign goods would decline and there would be a tendency for imports to increase. This would in turn force the rate of exchange, in the absence of counteracting factors, toward the gold export point, and if gold actually were to

flow out, it would tend to restrict the extension of credit and to raise the price of credit. Depression in the export industries and in industries facing increased competition from imports could not fail to have a depressing influence on the whole industrial fabric.

In other words, the result of stabilization at too high a point is at best a gradual retarding of business activity, made more difficult by rigid wage and interest rates; at the worst it means depression and possible abandonment of the gold standard.

The impact of these forces was softened for the moment and the seriousness of the situation in Great Britain obscured by the operation of a number of factors, one of which requires special mention, for it was a plague to all nations. This was the growth in the volume of liquid funds seeking profitable short-time investments in one market or another as the rate of interest might swing up or down, or looking for an opportunity to speculate in exchange rates, or more frequently seeking safety, without much regard to profit, in the market which seemed for the moment to be most stable. The return to the gold parity had induced in financial circles an unwarranted belief in the stability of London, and a large volume of liquid capital had been transmitted there for shelter and temporary investment. This obscured the real pressure on the exchanges, because the mechanics of transferring these funds to London served to counteract the tendency of changes in the currents of commerce to send the rate to the gold export point. They were, however, a constant source of weakness, for at the first hint of financial difficulties they would be (as later they actually were) transferred to another haven.

In summary we may say that the attempt to restore the gold standard at a mint par of exchange substantially above its purchasing power parity subjected the exchanges to a continual strain and in addition exerted a depressing influence, not only on the export industries, but on all other industries as well. An unstable equilibrium in the balance of payments was achieved largely by an influx of short-time investments, ready to move out at the first sign of danger and not counterbalanced by equivalent short-term investments by London in markets whence they could be repatriated. London did, it is true, have short-term investments in large quantities, but when she needed them they were "frozen" in Germany, Austria, and elsewhere.

When therefore the European crisis was precipitated in the spring

of 1931 by the failure of the Credit Anstalt of Vienna, and panic spread from Austria to Germany and then to Great Britain, gold began to be withdrawn in large quantities as panic-stricken investors transferred their funds to more stable markets. Despite active intervention in the exchange market, facilitated by large loans from the United States and France, it was impossible to hold on longer to the gold standard. In September, 1931, Great Britain was again on an irredeemable paper standard.

The gold exchange standard. Instead of returning to the gold standard in the period following the First World War a number of countries adopted the gold exchange standard. As explained in Chapter XIV this was a system by which certain countries made their money redeemable, not in gold, but in exchange on a country having the gold standard. The reason for this was a widespread fear of an impending gold scarcity, stimulated by pessimistic forecasts regarding the probable output of gold.

The basis of this fear lay in the fact that, despite a rise in the general price level, the price of gold in terms of the United States dollar (and later of the pound sterling) was allowed to remain at its 1914 level. Now the production of gold, like that of any other commodity, is affected by the relation of costs of production to the market price; and since costs of mining gold had increased while its price remained constant, gold mining became a less profitable venture than it formerly was. This condition was altered when the sterling price of gold was raised in 1931, and the dollar price in 1933—changes which caused the output of gold mines to increase in a spectacular way. But in the mid 1920's the fear of the scarcity of gold was so great that various devices were introduced to obtain "savings in the monetary use of gold."

Among these none was more detrimental to the success of the reconstructed system than the gold exchange standard.[1] This use by some countries of gold-convertible exchanges of another country as monetary reserves (instead of using gold itself) was harmful in effect because of the failure to work out a method of operating this modified form of the gold standard.

What might be called the rules of the game of the true gold standard had never been explicitly formulated. The system had

[1] For a detailed discussion of the shortcomings of the gold exchange standard, see Heilperin, M. A., *International Monetary Economics*. New York: Longman's, Green and Company, 1939, pp. 210–18.

functioned so well for many years that there was no need to codify specific rules of procedure; the nations worked together in the monetary field without them. But when this automatic collaboration was brought to an end by the many upheavals which had taken place, explicit agreements became necessary to give effect to the new exchange standard. They were not easy to reach, and actually no definite arrangements were ever made to insure the smooth functioning of the new system. The gold exchange standard soon began to crumble.

In order to assimilate intergovernmental war debts and reparations, and to develop new commercial and financial channels, two prerequisites were essential; namely, a large degree of freedom in the economic system, and the prevalence of confidence. Actually both were absent. Instead, the tide of economic nationalism ran high; tariff walls grew and multiplied; capital movements were haphazard and unstable; and movements of short-term balances, very erratic. Under the impact of the great depression which started in 1929, and of the growing protectionism, world trade declined considerably in volume, and its money value was further reduced by the acute price deflation of 1930 and the following years, so that trade balances proved inadequate to provide the means for settling international financial obligations.

During the international financial crisis which resulted, Great Britain was forced off the gold standard in 1931; many other countries either preceded or followed her; the United States gave up its gold parity in the spring of 1933, and the remnants of the gold bloc gave up by 1936. When the dissolution of the gold standard became complete, no basis remained for the operation of the gold exchange system in the international arena.

Stabilization of exchange rates: Flexible exchanges. The abandonment of the gold standard by the leading commercial nations, with its inevitable effect on the gold exchange standard, forced a consideration of practical exchange policy. It was argued by monetary theorists that exchange rates cannot be stable unless both internal and external price levels are stable. Since a country cannot control the external price level, it has a choice between allowing the exchange rate to fluctuate or, in order to keep the rate stable, allowing its price level to fluctuate in harmony with changes in the external price level.

As between these two alternatives, it was urged that national wel-

fare was best served by stabilizing the domestic price level and allowing exchange rates to fluctuate freely, seeking their own level. In accordance with the theory of the purchasing power parity, it was assumed that the level of exchange rates would be in the neighborhood of the purchasing power par. The conclusion drawn was in favor of a stable internal price level and flexible exchange rates. Nations accepting this conclusion were logically committed to a policy of abstaining from any international agreements designed to stabilize the exchange market.

Injurious effects of flexible exchange rates. Experience has not borne out the belief that exchange rates will always follow price changes without fluctuations so wide as to be disturbing to a nation's economy. In an earlier section we have seen how during a period of inflation the exchange rate moves more rapidly than the price index, anticipating the more slowly moving price level. Furthermore the fact that exchanges are not fixed by means of a mechanism such as was provided by the gold standard gives much encouragement to speculation and to upsetting movements of short-term funds from country to country. Under the gold standard, transactions in bank balances had a stabilizing effect; in the absence of exchange stability, their effect is disturbing. This was proved by the world's experience during the decade of the 1930's.

In addition, the gold standard involved a certain measure of "international discipline." Under that system price movements taking place in the various countries were linked together, in the sense that no country could remain within the system and have at home an independent policy of inflation or deflation.

The abandonment of the gold standard gave each nation more freedom of economic action, but such freedom could only be won at a price. Under flexible exchanges the price is paid in terms of general international instability. Eventually instability leads to a collapse of confidence, and this expresses itself in massive capital flight; *i.e.,* in large and sudden transfers of funds from one currency into another. The world had in the 1930's a great deal of experience with capital flights, and it was found that no country could long endure them without such serious consequences that the government would take steps to prevent them. It was these capital flights which finally discredited the doctrine of flexible exchanges.

Exchange control. While exchange restrictions were not unknown before the onset of the depression of the 1930's, their extension during that decade was at the same time the result of the depression and a cause of its continuance. A surprisingly large number of countries in Europe and South America adopted, during these years, more or less elaborate systems of control, Germany being an extreme example. Devised, as we shall see, as protective measures in troubled times, these controls fitted rather perfectly either into plans of managed currency or into more comprehensive plans of national control of economic life. They also became a helpful instrument of policy for countries, like Germany, whose ultimate aim was war.

To understand why so many countries have resorted to restrictive controls, it is necessary to recall that, in the period following the First World War, a precarious balance in payments was achieved in many of them only by large loans from the United States, that their gold reserves were dangerously small and their monetary systems easily thrown into confusion, and that many of them were over-burdened with interest payments, short-time investments, and debts for past importation of goods.

The crisis in the United States put a complete stop to the floating of international securities in the American market, thus placing a strain on the balance of payments of all countries which used exchange acquired by these loans to meet current indebtedness. At the same time all countries found the world demand for their products declining, and the world prices of their products—particularly of agricultural, mineral, and other staple goods—decreasing rapidly.

The danger to the exchanges and to internal monetary and banking structures seemed in many countries critical enough to warrant drastic measures. The most obvious move was to stop the transfer of the payment of debts owed abroad. The Hoover Moratorium of June, 1931, on the payment of interest and principal on intergovernment debts and reparations set a precedent which was willingly followed by many countries.

Moratoria, expanded to postpone the payment of all indebtedness whether arising from bonded indebtedness, short-term investments, or imports, served to prevent the obligations of a more optimistic past from exerting a disastrous influence on the balance of payments, but in many countries equal concern was felt lest

current international transactions have a similar influence on the balance of payments, and some countries were anxious to acquire a favorable balance from current transactions, which would enable them gradually to liquidate their unpaid foreign debts.

To attempt to secure a balanced account or a favorable balance through an increase in exports seemed futile; the only other method was a reduction in imports. In the long run, this is not only equally futile, since one country's imports are the exports of some other country, but it is positively harmful, because of the influence on the volume of world trade of a series of import restrictions. Yet it was an expedient which might secure results for a few weeks or a few months before other countries adopted similar measures.

Effective command of the situation demanded a concentration of exchange control in the hands of the government, for thus only could undesired fluctuations in the rates be prevented, and only in this way could all exchange acquired by exports be made available both for the payment of imports and for the liquidation of such portions of past indebtedness as seemed expedient. This type of restrictive control involves usually these three things: (1) the surrender of all foreign exchange by exporters and others to some representative of the government, (2) a system of import licenses, issued according to the availability of exchange and sometimes according to the importance of the import to national life, and (3) an official rate of exchange set by the exchange authorities. A considerable, though decreasing, degree of difficulty has been experienced by most countries which have attempted to monopolize exchange transactions in this way, for coincidental with the expansion of official control came the establishment and spread of what were known as the black bourses—subterranean markets for "bootleg" exchange.

Official action to balance imports and exports with each country. In all countries which exercise strict control of the type just described, great importance is attached to the balance between exports and imports of goods, because of the fear of the influence which an unfavorable balance of trade might exert on the exchanges. In some countries, indeed, the novel doctrine has been evolved that a balance in exports and imports should be secured with each separate country. It has a specious appearance of fairness, and a surprisingly large number have been seduced by it. For example, Germany informed America that unless Americans bought more from her she

would reduce her American imports, and she actually did so, bringing her American exports and imports into balance.

While this type of balance may offer some practical advantages to countries which have set up exchange clearing arrangements, it rests otherwise on a complete misunderstanding of the principles of international trade, according to which it is the total balance of payments—not the separate trade balances with particular nations—that affects exchange rates and the monetary situation. The final result of a forced equilibrium of exports and imports with particular nations can be only a decrease in the total volume of world trade, a decline in interregional division of labor, and a shrinking in world income. To justify this policy as a weapon of commercial diplomacy is to mistake the bases of world trade; exports and imports exist and increase because of the mutual advantages secured from the trade and not because of bludgeoning methods used by one country towards another.

Exchange clearing systems. Exchange clearing systems represent in origin a device whereby one country may secure payment for exports to another country, or for interest which is due but withheld. As a weapon of this sort it is available only to a country which has an unfavorable balance of trade with another country. In actual practice the clearing system may have been established by mutual agreement, but behind the appearance of an amicable settlement there is the possibility of forcible collection.

To illustrate this let us take the trade of Hungary and Switzerland. Switzerland was importing more from Hungary than she exported to her and was paying for her imports, whereas the Hungarian government decreed that payment for Hungarian imports must be made in national currency into a special account with the Hungarian National Bank. Assuming that the Swiss exporters were not content to have "blocked" accounts in the Hungarian National Bank, the reasonable solution was either for Switzerland to cease exporting or to impound in Switzerland the money Swiss importers owed Hungarian exporters, make a deduction sufficient to cover payment to the Swiss exporters, and remit the balance to Hungary.

As a matter of fact Hungary was induced to see the reasonableness of this, and in November, 1931, an agreement was signed whereby payment for Hungarian exports to Switzerland was to be made in francs by the Swiss importer into a special account in the Swiss National Bank, and payment for Swiss exports to Hungary

was to be made in pengös by the Hungarian importer into the clearing account in Hungary. Thus the payments of Swiss importers made funds available in Switzerland for payments to Swiss exporters, and the same was true in Hungary. The arrangement included practical details for balancing the two clearing accounts, which need not concern us here.

Undoubtedly the system has practical merits in trading with a country of weak financial standing. It gives the countries effective control of their trade balances; it insures payment to exporters of both countries; it removes the exporters and importers entirely from the exchange markets, leaving the matter of transfer to governmental officials; and it is probably true that it removes some of the incentive to erect additional tariff barriers, exchange restrictions, and the like.

Because of these practical advantages many of the countries which have adopted stringent restrictive measures of control over the foreign exchanges have turned to exchange clearing systems. As a part of a permanent trade policy it is of more than doubtful wisdom, unless the goal be the control of the character of international trade and not an expansion in its volume.

Measures to revive international monetary coöperation. The "sterling area." [1] At a time when certain countries were perfecting more and more effective instruments of monetary insulation and nationalistic control, other countries were striving to restore a certain minimum of international monetary coöperation. The development after 1931 of a "sterling area" based on Great Britain was an important development in this direction.

This group included the members of the British Commonwealth (with the exception of Canada) and several other countries, such as Portugal and Scandinavia, which regarded the stability of sterling exchange as essential to their foreign trade. After the outbreak of war, the sterling area was reduced to members of the British Commonwealth and developed a strict centralization of foreign exchange dealings in London, as an aspect of the wartime policy of exchange control. Prior to the war, however, members of the sterling area did not exercise exchange control, but managed nevertheless to maintain stable sterling exchange rates. Had France and other remaining gold standard countries joined the sterling

[1] Cf. League of Nations, *International Currency Experience*. New York: The Columbia University Press, 1944.

area, as had frequently been proposed, we might have experienced an early reconstruction of monetary collaboration over a large part of the world.

But this did not happen, and when the United States abandoned the gold standard in 1933, uncertainties as to the future gold value of the dollar further disrupted monetary relations throughout the world. The Gold Reserve Act of 1934, which set the price of gold at $35 an ounce, 59 per cent of the old parity, exerted a stabilizing effect on the exchange market. After 1934 the dollar price of gold remained stationary, and until the outbreak of the Second World War, the dollar-sterling rate also remained almost stable.

The exchange equalization system. Another instrument of monetary control was introduced in England in 1932 and in subsequent years was adopted by several other countries, including the United States, France, etc. It was called "The Exchange Equalisation Account" in England and "The Stabilization Fund" elsewhere.

The purpose was to provide means for a systematic governmental intervention in the foreign exchange markets with a view to eliminating fortuitous exchange fluctuations, while allowing exchange adjustments which have their roots in basic economic conditions to take place. Those administering the stabilization fund recognized the existence of economic nationalism and arrived at the modest objective of orderly exchange adjustments when the situation called for them. This represented a sort of compromise between the complete exchange stability of the gold standard and the undesired exchange flexibility discussed in an earlier section of this chapter, and it was therefore a significant step in the direction of new forms of international monetary organization.

It is interesting to contrast the operation of an exchange stabilization fund and that of centralized exchange control. The former used the market mechanism, with the Fund intervening as buyer or as seller of sufficiently large amounts of foreign exchange (or of gold) to achieve the desired objective. The latter suppressed the market and put in its place a governmental monopoly of dealing in foreign balances and notes. Exchange funds are compatible with a market economy based on free enterprise, whereas exchange control typifies state socialism.

The Tripartite Agreement. When by 1936 the economic and financial situation in France so deteriorated as to make her going off the gold standard inevitable, another important event took

place in the field of international monetary relations. In order to forestall a new period of competitive devaluations, the British, American, and newly established French Exchange Funds concluded an agreement which was to govern relations between them.

The so-called Tripartite Agreement represented a compromise "whereby the exchange funds of countries with fixed [1] and flexible [2] gold prices could have access to each other's markets and could coöperate in the management of gold shipments in both directions. The prices governing official transactions in the countries with flexible gold prices were held stable, for a period of time long enough to enable the authorities of countries with fixed gold prices to complete the conversion operation without risk. The importance of this agreement was in the antecedent understandings concerning the exchange rates on which gold prices were based, and its effect was to allow the various exchange funds, central banks, and treasuries to operate an international gold settlement system that was an amalgam of conflicting techniques and principles." [3]

The Tripartite Agreement was soon joined by Switzerland, Belgium and the Netherlands. Combined with the sterling area, it provided the nucleus of what might have eventually become a new international monetary system. The Second World War broke out before any further development had time to take place.

By 1943, in anticipation of the postwar reconstruction, discussions had begun which aimed at the establishment of a monetary system for the world emerging from war. These discussions led eventually to the Bretton Woods Agreements of July, 1944, and to the establishment of the International Monetary Fund. The Bretton Woods Conference also established the International Bank for Reconstruction and Development described in a previous chapter.[4] By adopting the Bretton Woods Act, Congress made the United States a participator in both institutions.

The International Monetary Fund: Objective. This institution is the outgrowth of the experiences with stabilization funds referred to above and represents an attempt to organize stable monetary relations in a world disrupted by war and divided by highly nationalistic policies. However desirable in the abstract a return to the gold

[1] In this instance the United States.

[2] Great Britain and France.

[3] League of Nations, *International Currency Experience*. New York: The Columbia University Press, 1944, p. 147.

[4] See Chapter xxvii.

standard might be, the system in its pure form appeared in 1943 not to be practicable. There may, however, be a gradual return to free international payments, the maintenance of a considerable degree of exchange stability, and the establishment in the monetary field of international machinery to take the place of isolated actions on the part of separate national governments.

The Fund aims at all these objectives. It is a coöperative endeavor to grapple with problems which, after the First World War, failed to be so handled, although the environment at that time was more favorable. The future alone will tell whether these new endeavors will prove successful. Following is a brief survey of the principal provisions of the Fund's charter.[1]

The organization of the Fund. The membership of the Fund consists of the governments of various countries. The Fund will deal only with the treasuries, central banks, stabilization funds, or similar agencies of the member states. The resources of the Fund consist of subscriptions made by members according to a schedule agreed to at Bretton Woods. The size of the Fund will depend on the actual membership; countries represented at Bretton Woods, if they all joined, would establish a fund with resources of 8.8 billion dollars, of which the United States would contribute 2,750 millions and Great Britain 1,300 millions.

The contribution of each country is called its *quota* and occupies an important place in the future relations between that country and the Fund. Of this quota, the member pays 25 per cent in gold, the rest in its own currency, special exceptions being provided for countries whose gold reserves have been depleted by the war. Each member must define its currency in terms of gold, and should this valuation be subsequently altered, the member must pay into the Fund enough of its currency to maintain the gold value of the total amount held by the Fund.

Operations of the Fund. The main purpose of the Fund is to assist its members in meeting temporary deficits in their balances of payments. It is not its purpose to provide members with long-term capital, and members are warned not to use in an improper way the facilities of the Fund. Penalties are provided in case of abuse, including the suspension of membership.

The Fund sells foreign exchange to members against payment in

[1] Cf. Heilperin, M. A., *International Monetary Reconstruction.* New York: American Enterprise Association, 1945, pp. 23–53.

the member's own currency. A member, unless the Fund rules otherwise, is entitled to obtain foreign exchange up to the total amount of its quota, but not more than 25 per cent of that amount in any one year. As the member country's foreign exchange position improves, it is expected to repurchase from the Fund appropriate amounts of its currency, paying in gold, dollars, or other currencies. In this manner it is proposed to safeguard the character of the Fund as a revolving fund so that the institution can assist one country after another as they encounter balance-of-payments difficulties.

The Fund will not interfere directly in the operations of foreign exchange markets, but through its own operations it is expected to remove from the markets exceptional pressures which might gravely upset the equilibrium of currently quoted exchange rates. The foreign exchange markets will continue to function in the manner previously described. The new institution will enable central banks which have depleted reserves because of an outflow of gold or foreign exchange to obtain additional means of payment from the Fund. The Fund, it is hoped, will therefore become a sort of "reserve bank" for the central banks and treasuries of the member countries.

Exchange stability: General provisions. It is stated in the Fund's charter that "the par value of the currency of each member shall be expressed in terms of gold as a common denominator or in terms of the United States dollar of the weight and fineness in effect on July 1, 1944." Provision is made for the possibility of actual quotations oscillating within narrow limits around that parity—a sort of substitute for the gold standard's system of "gold points."

The Fund, on the other hand, contains elaborate provisions for alterations in a par of exchange when necessary "to correct a fundamental disequilibrium" in any member's balance-of-payments situation. A practical example of "fundamental disequilibrium" might be the member country's inability to maintain its balance of payments in a condition of long-run equilibrium without either a very disruptive domestic deflation or the adoption of strict controls on its foreign trade and payments.

When for this or other reasons, changes in parities are proposed by the interested member country, the Fund is to adopt the following attitudes:

"If the proposed change, together with all previous changes, whether increases or decreases,

(i) does not exceed ten per cent of the initial par value, the Fund shall raise no objection;

(ii) does not exceed a further ten per cent of initial par value, the Fund may either concur or object, but shall declare its attitude within seventy-two hours if the member so requests;

(iii) is not within (i) or (ii) above, the Fund may either concur or object, but shall be entitled to a longer period in which to declare its attitude."

Thus we have two 10 per cent "zones of tolerance" for parity changes; if the proposed change exceeds the limits of tolerance, the matter is viewed as a problem of international concern and becomes the object of negotiations between the country in question and the management of the fund. Should a member country change the par of its currency in disregard of objections raised by the Fund, "the member shall be ineligible to use the resources of the Fund unless the Fund otherwise determines." In case of a serious conflict between a member and the Fund, the former may be required to relinquish its membership. Thus the "Bretton Woods System" provides a framework for amicable settlement of matters which used to be the object of separate national decisions. In the unsettled world in which we live, these provisions may possibly help to avoid a repetition of the 1930's.

Initial parities. The determination of initial parities is a very delicate matter, in view of the impossibility of accurate determination. Since a "good" par is one that can be maintained without domestic hardships and international tensions, the only effective way of determining it will be by the process of trial and error.

The world price of gold. At the time of the inauguration of the Fund, the world price of gold was $35 an ounce. Should this price prove to be inadequate to meet the monetary needs of the world, it can be modified by means of a uniform change in par values agreed to by the majority of the voting power in the Fund. It is provided, however, that the change must be approved by every member holding 10 per cent or more of the total of the quotas, which includes the United States.

Exchange restrictions and freedom of international payments. The charter of the Fund provides that, after a transitional period of five years, all international payments connected with current commercial and financial transactions shall be free from governmental restrictions. This includes income payments on foreign in-

vestments and moderate amounts for amortization of loans or for depreciation of direct investment.

Movements of long-term capital, on the other hand, and purely financial transfers of short-term funds may remain subject to controls: "Members may exercise such controls as are necessary to regulate international capital movements, but no member may exercise these controls in a manner which will restrict payments for current transactions or which will unduly delay transfers of funds in settlement of commitments."

Furthermore, a member of the Fund must use exchange restrictions in case of capital flights rather than resort to the facilities of the Fund, since it is expressly stated in the charter of the Fund that "a member may not make net use of the Fund's resources to meet a large or sustained outflow of capital." This decision was a compromise between America's demand for entirely free international payments and the insistence of practically all other countries on the maintenance of some degree of exchange control.

Recognizing the close connection between technical monetary relations and the larger problems of international trade, the Bretton Woods Conference adopted a resolution calling upon the members of the Fund to coöperate in removing restrictions on international commerce and promoting the orderly operation of the trade and money markets.

CHAPTER XXX

Economic Nationalism and the Protective Tariff

International division of labor. That division of labor between individuals promotes individual well-being in the economic sense has been fully demonstrated, by experience as well as logic. The advantages to be gained by division of labor between different regions or nations are of the same order and equally demonstrable. This subject has already been developed in connection with our study of interregional trade,[1] where it was made clear that any nation or region obtains for its people the highest possible standard of living (other things being equal) when it devotes its industrial efforts to those lines of production in which it has a comparative advantage and depends for other products upon the industrial efforts of other parts of the world.

National self-sufficiency. In spite of these clear truths, the peoples of the world have never accepted fully as a matter of practical policy the corollary of freedom of international trade.

Considerations of various sorts have led to governmental restrictions upon the trade between nations. Sanitary and moral considerations demand restrictions upon the import of certain products, such as narcotic drugs. Tariff duties are laid upon imports and exports in order to bring in public revenue. National preparedness for war is sometimes believed to require the fostering of certain industries within the nation, even though this may be accomplished only by placing obstacles upon the import of foreign-made products. In recent times certain nations have believed that the protection or control of their monetary systems required them to place restrictions upon imports.

But most potent of all the influences leading to restriction of the freedom of international trade has been a popular belief, widespread and persistent, that by checking imports, a nation may enhance its economic well-being in terms of greater wealth and a higher standard of living.

[1] See Chapter XXVII.

In consequence of all these considerations—and especially the last—there has developed an ideal of national economic organization which is the antithesis of international division of labor and which may be called "national self-sufficiency," or "economic nationalism." This ideal requires that a nation be, so far as possible, free from reliance for important products upon the industry of other peoples. For its accomplishment there is required obviously severe restriction of imports from abroad; indeed a tendency to regard all imports as a menace to national prosperity is evident.

While the theory of nationalism does not necessarily imply a forcing of exports, the fact is that those who hold to this ideal generally regard exports with a favor no less than their disapproval of imports. Imports are bad; exports are good; the former impoverish, the latter enrich the nation.

This ideal of industrial nationalism was a cardinal principle in the philosophy of the mercantilists, and its accomplishment was consciously sought in the practical policies of the leading nations of Europe in the seventeenth and eighteenth centuries. Its theoretical foundation was demolished by Adam Smith. It later acquired a new lease on life through the teachings of the German economists of the so-called "historical school," predominantly Friedrich List, and Henry C. Carey in the United States. It is an article of faith of the protectionist movement which has flourished in many nations for many generations.

At the moment, the idea of industrial nationalism has gained a hold upon popular thinking stronger perhaps than at any time since the days of the mercantilists. It is closely interwoven with commonly held doctrines of the causes and cure of depression and is apt to be an essential part of current conceptions of the planned economy. It must finally not be forgotten that among the most persistent apostles of national self-sufficiency are many whose own self-interest would be promoted either by the restriction of imports generally or—what is more to the point—of the particular good which they produce in competition with foreign makers.

Effects of the protective tariff. For the accomplishment of the ends of national self-sufficiency a variety of means is at hand. Embargoes, limitation of permitted imports by national quotas, restrictions upon foreign exchange, depreciation of the national monetary unit—all these are devices which have been employed, never

perhaps so widely or vigorously as during the past generation. But the chief instrument for the restriction of imports is the protective tariff, and the whole subject of economic nationalism may be elucidated in large measure by analysis of the problem of protectionism.

For our excursion into this problem the principles of international trade, as developed in an earlier chapter,[1] furnish the starting point. We know that, in the absence of restrictions upon its foreign trade, the people of each nation tend to buy goods in the cheapest market and sell in the dearest market and that this leads them to devote their productive resources and energies to those lines in which their nation has comparative advantage, as evidenced by the ability to produce at comparatively low cost. At the same time they rely upon other nations for those goods in whose production they are at comparative disadvantage. Thus develop international trade and international division of labor.

International trade may be checked by means of duties on imports.[2] A duty high enough to be protective may diminish or entirely stop the import of any chosen commodity.[3] Carried to its extreme limit, the protective tariff might be used to prevent virtually all imports. Whether applied moderately or to the limit, the protective tariff has the effect of reducing imports and building up corresponding industries at home. It thus restricts international trade, promotes "diversity of industry" in the home country, and diminishes reliance upon the goods of other nations. National self-sufficiency gains at the expense of international division of labor.

As to the truth of the foregoing statements no question can be entertained by anyone who is familiar with the elementary principles of economics. The controversy over protection starts from this point. Granting the respective effects of freedom of trade and protection as we have stated them, the "protectionist" asserts that it is to the economic advantage of a nation thus to increase diversity of industry and national self-sufficiency; the "free trader" [4]

[1] Chapter XXVII.

[2] Since they seldom enter into the controversy—never in the United States—no consideration need be given in this discussion to duties on exports.

[3] See Chapter XXIV, pages 473-77.

[4] The reader should perhaps be reminded that "free trade," as the term is used, does not mean no tariff; it means no protection. The free trader has no quarrel with the revenue tariff; he is opposed only to protection. There is no country in the world entirely without a tariff and no appreciable sentiment in favor of the abolition of all tariffs.

says that this means an economic loss and that the nation gains most when it enjoys all the advantages of international division of labor. Thus the issue is sharply drawn.

Protection and national self-sufficiency: The case for protection. The advantages of trade, as already developed, are so evident and so important that we are justified in placing the burden of proof upon the protectionist, who claims that there is an economic gain from putting artificial restriction upon a nation's foreign trade. Let us proceed then to examine the principal arguments advanced in support of the policy of protection.

a. The favorable balance of trade argument. Some of the most popular arguments for protection are such naïve applications of old mercantilist fallacies that they may be dismissed with nothing more than brief reference to the established principles of trade. Such is the argument that a nation gains in its foreign trade only from its exports, that imports are a national loss, and that the tariff is therefore beneficial through its effect in checking imports. We have already learned that, so far from exports being superior, it is imports which are the real goal of international trade; we have learned also that a country's balance of payments must in the long run be in substantial equilibrium, and that therefore anything which reduces imports must eventually also reduce exports. It is neither advantageous nor possible by means of a tariff to make imports decline relatively to exports.

b. The gold argument. Similarly naïve is the argument that a favorable balance of trade, created by the tariff, causes gold to flow into the country and is therefore advantageous. This is simply a corollary of the preceding argument and rests upon other mercantilist fallacies with which the reader is acquainted. We know that it is not the balance of trade (*i.e.,* merchandise exports and imports), but the balance of payments (including all debits and credits), which determines the inflow or outflow of gold; and we are also aware that it is neither possible nor advantageous for a country to pile up gold as the result of artificial restrictions upon its foreign trade.

But there are those who are not content with this summary disposal of cherished beliefs. Is it not true, they say, that the rule for amassing wealth is to sell much and buy little? Is it not thus that the individual businessman accumulates money from the profits of his business? And must not the same rule hold for national gain

from trade? At the risk of adding proof of the obvious, we may give some further attention to these mercantilist notions.

The wealth of the people of the United States in 1925 was estimated at about 355 billion dollars. At that time the nation's total stock of gold coin and bullion was about 4½ billions, or about 1⅓ per cent of the national wealth. In 1880 the national wealth was, according to estimates of the United States Census, about 44 billions. The stock of gold coin and bullion was then 352 millions. In this period of 45 years therefore the national wealth was increased by 311 billions, while slightly over 4 billions were added to the stock of gold. It is certain enough that this great gain in national wealth was not due to accumulation of gold from foreign trade. The growth of a people's wealth is not principally in the form of gold; it consists in the increased value of its lands and houses, its factories and railroads and ships, its homes, and countless other forms of wealth vastly more important than its stock of gold. And the gain from its foreign trade is not in the form of gold but in the useful commodities imported.

But what of the analogy of the businessman? The answer is that neither does he accumulate wealth in the form of money or by selling more than he buys. If, at the end of a profitable year in which a man has made gains of $50,000, we should ask him to show us his profits in concrete form, he would not point to an increase of his money on hand or his bank deposits. The probability is that these items are no larger than at the beginning of the year. His profits were not made by selling more than he bought. As he received money from the sale of goods, he paid it out—partly for current consumption, partly for additions to his factory and machinery, for lands, for improvements to his home, for a new automobile, for government bonds and corporation securities. Thus he will show concrete evidence of his year's accumulation. So it is with the nation. A nation, like an individual, grows wealthy, not by selling more than it buys, but by producing more than it consumes.

c. The added industry argument. It is argued that the new industries which admittedly may result from protection are an addition to the industrial resources of the nation and that therefore the tariff has increased the national wealth. Now the only way that the tariff can cause the establishment of a new industry is by making it too costly to buy a foreign product previously imported. Also we have learned that this diminution of our imports must be bal-

anced eventually by a corresponding decline of our exports. The balance of payments must maintain its equilibrium; if we will not buy of foreigners, they cannot buy of us. The result is that, while a new industry is arising, some old industry finds its foreign market declining and is forced to reduce its operations accordingly. The new industry is thus not an addition to the nation's industry but merely a diversion of capital and labor from one field to another.

Whether capital and labor already employed are shifted or new capital and labor are directed into the protected field, there is clearly no gain in total industry. Moreover, the diversion of industry is from the more profitable into the less profitable fields and so results in a net economic loss.

d. The wages argument: Increased demand for labor. Of the two forms taken by the wages argument for protection, the first is that the protective tariff, by increasing the total amount of industry in the country, adds to the demand for labor and so tends to increase wages. Protection is therefore of advantage, not merely to the entrepreneur class, but to all laborers, and so to at least the great majority of the people.

It is evident that this argument rests for its foundation on the added industry argument and falls with it. The mere diversion of industry, particularly when it is from more advantageous to less advantageous lines, can at best do nothing to increase the demand for labor or to raise wages.

Support of high wages. The wages argument is not always put so crudely as this. By American protectionists it is more often argued that, in the absence of the tariff to keep out cheap foreign goods, American manufacturers could not pay their present high scales of wages but would be compelled to lower wages or perhaps to go out of business entirely. Thus the tariff is in the interest of the workingman.

This argument overlooks the fact that many American industries seem able to get on well enough without protection. These are the industries naturally fitted to prosper in the American environment, the industries in which America has a comparative advantage. Without protection, they are nevertheless able to pay the prevailing high level of American wages; they pay as high wages as the protected industries. If any industry is really unable to pay the prevailing rate of American wages, it is because that industry is not adapted to our conditions. Its presence is due to the artificial stimu-

lus of the tariff, and it exists at the expense of some more favorable industry which could pay high wages without tariff assistance. Removal of the tariff might cause the artificially favored industry to disappear, but its loss would be offset by the gain in the naturally favored industry, and there would be no decline in wages.

Wages and labor cost. There is much loose talk about wages in this connection. Protectionists argue as though, simply because wages are higher here than elsewhere, it must cost more to produce anything in America. They express alarm at the prospect of competition with the products of "foreign pauper labor." But the fact is that in numerous fields of agriculture and manufacture, goods are produced more cheaply in America than anywhere else in the world, in spite of high wages. In these fields employers are able to pay high wages for high-grade labor, and that without making the cost of production high.

In international competition, as in industrial competition everywhere, the controlling factor is not high or low wages, but high or low *labor cost of production.* Wages, as such, tell us nothing. It is only when wages are divided by quantity produced, to give the labor cost per unit of product, that we have the key to the problem of relative advantage in international competition. Thus it is that high relative wages, when accompanied by still higher relative output, give low labor cost.

This is the condition in the great body of industries which are naturally adapted to American conditions. Here high wages are paid without the aid of protection. In general these are industries which either are favored by peculiar advantages of soil or climate or natural resources, such as cotton growing, wheat growing, copper mining, etc., or have the advantage of skillful use of labor-saving machinery, such as the manufacture of automobiles, sewing machines, farm machinery, etc.

Industries not so favored by nature, or which must lean heavily upon hand labor, do not flourish here. To them the high level of American wages is a handicap, which can be overcome only by a protective tariff which removes the competition of the cheaper foreign product. Protection is required therefore, not to maintain the high level of American wages, but to make possible the payment of American wages in industries not suited to American conditions.

e. Alleged danger of withdrawing protection. The wages argument is without doubt urged in good faith by many American

manufacturers who honestly believe that protection is necessary, not merely to their own profits, but also to the well-being of American labor. Being engaged in lines of industry which possibly could not exist without tariff protection, they believe that without protection they could not continue giving employment to their labor forces at the present scale of wages. They visualize the distress which would follow a material cut in wages or a shutdown of industrial plants. They fail to see that discontinuance of the artificially fostered industry would eventually open up other fields in which these laborers would find employment at American wages.

It is not to be denied that the shifting would impose inconvenience and loss upon the laborers and still more upon the capitalists. Industry that has become adapted to the artificial condition of protection cannot adjust itself to freedom of trade instantly or without some loss to all concerned. This is no argument for retaining permanently the artificial condition. It simply dictates caution and consideration in making the change.

The fact is that American industry would not be wrecked by a careful, gradual removal of protection. Some few industries would doubtless succumb. In most cases certain grades or lines would be discontinued while others advanced, the sort of adjustment which normally goes on as a matter of course in every progressive industry. For industry as a whole the final effect would without question be beneficial.

American tariff history shows protection surprisingly less potent, either for good or for evil, than both its friends and its enemies would have us believe. The course of American industry has not been vitally influenced by its imposition, nor would it now be revolutionized by its withdrawal.

f. Alleged saving in transportation costs. As an argument in favor of protection, attention is sometimes called to the supposed waste involved in sending raw materials to a foreign country, only to be returned to the country of origin in the form of manufactured products. An example is American cotton, much of which goes abroad to English cotton mills, and some of which finds its way back to America in fine British cotton fabrics. Here is said to be a wasteful double cost of transportation, which might be avoided by a tariff that would keep out the British product and compel its manufacture in America.

The fallacy here is obvious. There is no loss in the operation in

question; otherwise it would not occur. Transportation costs are no worse than other costs. The double transportation of cotton across the Atlantic is undertaken because that is the way to get the product to the American consumer at the least total cost. Forbidding this method by a tariff would save this transportation cost, to be sure, but only at the expense of a still greater cost of manufacture.

The *reductio ad absurdum* of this argument appears if we imagine ourselves trying to apply it universally. Practically all domestic industry involves such double costs of transportation. Wheat is raised on Nebraska farms, shipped to Minneapolis, ground into flour, and returned to the Nebraska farms. Iron and copper and coal are brought from Minnesota and West Virginia to New England, and the finished steel and brass goods are shipped back to these states. The sole reason for all this double transportation is to minimize total costs of production. Its avoidance would be possible only at the cost of abandoning territorial division of labor and putting the human race back in the Middle Ages economically.

g. Alleged need to meet the tariffs of other nations. There are those who are ready to admit the advantages of universal freedom of trade between nations, but who yet feel that, in a world in which protective tariffs are common, no one nation can afford to permit free entry of foreign goods, nor can one nation safely remove its tariff barriers unless other nations do the same.

This misgiving overlooks the experience of Great Britain, which, during half a century in which she led the world in an amazing industrial development, maintained freedom of trade regardless of the tariffs imposed against her products by other nations. Such an attitude overlooks also the fundamental principle of trade. It is true that our own industries and our economic welfare are injured by the tariff barriers of other nations, which deprive us of some measure of the benefits of international trade. But the whole conclusion of our analysis has shown that our national welfare is injured by the curtailment of our foreign trade occasioned by our own tariff. Freeing our trade from the shackles both of our own tariff and the tariffs of foreign nations would be doubly advantageous. But if we cannot secure the removal of foreign tariffs, that is no reason why we should forego the gain to be derived from removal of our own tariff.

The trouble with reciprocity agreements, tariff treaties, and the like is that they are usually conceived in the spirit of mercantilism,

which regards exports as a national blessing and imports as a national curse. There follows a sort of inverted horse-trading negotiation, in which each nation seeks to give the other the utmost of its products while taking as little as possible in return.

Once the notion is reversed and imports are recognized as the true goal of foreign trade, with exports desirable only as the means whereby imports may be obtained, the whole spirit of mutual tariff concessions is changed. No nation need grudge other nations whatever advantage they may derive from relaxation of its own tariff. If concessions in foreign tariffs may be obtained in return for reduction of our own tariffs, this is all to the good; but it does not follow that our own relinquishment of trade obstructions should be abandoned because other nations do not see fit to act likewise.

h. Alleged gain from injuring other nations. Mercantilist notions sometimes run to the extreme belief that whatever injures the industry and trade of another nation must be beneficial to us. It follows that one of the chief services of our tariff policy should be to injure the economy of foreign nations by excluding their goods from our markets. That can of course be done.

We have seen that in that case the principal injury is to ourselves. We suffer also from whatever injury our tariff policy may inflict upon other nations. The whole world is bound together by the universal principle of division of labor. Just in proportion as the industries of other nations are efficient and productive will they be able to serve our needs by enabling us to obtain through exchange with them certain products at less cost than if we had to make them ourselves. In general each nation stands to gain economically, not from the adversity, but from the prosperity of other nations.

Considerations against protection. The case for free trade consists principally in demonstrating the fallacy of the arguments for protection. There are, however, some important positive considerations on the free trade side of the argument which serve to throw further light upon the question.

What is the territorial unit for protection? It is an interesting characteristic of the foregoing arguments for protection that each assumes that the tariff barrier will follow national boundary lines. Yet the arguments themselves are purely economic; they contain nothing to show that protection would not be exactly as beneficial if directed against a competing region within the same nation. We are tempted to ask why the New England textile mills should be

denied protection against the rising industry of the South, and vice versa, or why the automobile industry in various states should not have protection against Michigan.

Historically many of the first tariffs were imposed by towns or other minute political subdivisions. They proved such a burden on trade and such a drag on economic progress that they were swept away. In like manner it came to be recognized that the tariff systems of some score or more of little states in central Europe were proving fatal to the economic development of that region, and the German *Zollverein*—customs union—swept them all away and opened the path to a later political union and to the marvellous industrial development of the German Empire. Originally the American states had each its own tariff system directed against the others. They had to give way in the interest of the economic welfare of the whole region. Few today would seriously propose to retrace any of these steps already taken in the direction of freedom of interregional trade. No valid argument exists to show that still further progress would not follow the removal of the remaining barriers to the trade between nations.

Benefits of foreign competition. It is a fallacy to regard foreign competition as especially heinous. Manufacturers naturally dislike competition. They seek to avoid it whenever they can. But competition is in the interest of the consumers, for whom industry in the last analysis exists. And the competition of foreign producers is just as wholesome as that which is nearer at hand. If the home manufacturer cannot meet the foreign competition, then the home consumers are entitled to the less costly product. If the home manufacturer, spurred on by the foreign competition, does succeed in meeting it and holding his market, then again the consumers are benefited.

Moreover the nation gains from the increased efficiency of an industry which might otherwise have lagged behind. There are instances of American industries which, if pushed to do their utmost, could have competed on even terms with the best of the foreign producers, but which, sheltered by the American tariff wall, have grown lax and fallen behind the industrial procession.

Conclusion. Analysis of the arguments for and against protection thus appears to demonstrate that industrial self-sufficiency secured by means of the protective tariff, so far from enhancing the national welfare, is a cause of economic loss. No gain can be shown

to offset the demonstrated advantages of international division of labor. Protection is an economic burden, a cause of loss to the people of the nation which employs it. The economic advantage is all on the side of freedom of international trade.

It is probably not necessary to remind the reader that these conclusions do not imply that any nation—least of all the United States—could instantaneously abolish its entire tariff system without some unhappy consequences. Tariff systems have grown up during a long period, and the industry of each nation has adapted itself to the tariff situation as it has been and is. In America a system of very high tariff duties upon most of the important products of manufacture and upon some raw materials has been in effect for three-quarters of a century. There have in this period been variations in the level of protection, but never anything approaching free trade. American industry has grown up in this environment, to which it has had constantly to adapt itself.

To change this condition, suddenly and without notice, by going over to the free entry of all foreign-made goods would throw industry into confusion and do serious injury to the economic welfare of the nation. From this proposition—which must be well-nigh self-evident—the conclusion is, not that free trade is not the ideal or that protection must be retained, but rather that the change from a condition of protectionism to the healthier condition of free trade must be made gradually and with due notice, to the end that the costs of readjustment to industry and the inconvenience to the people be made as small as possible.

Temporary protection. Thus far we have been considering the question of protectionism from the viewpoint of the relative merits of international division of labor and national self-sufficiency, the protective tariff serving as a permanent device for promoting the latter ideal. In contrast we must now observe the possibility that the tariff may be called upon from time to time to serve certain temporary objectives not involving any permanent artificial interference with the freedom of international trade. The classical example of this application of protection is the use of the protective tariff on behalf of young industries.

Infant industry protection. Use of the protective tariff has been frequently urged and occasionally employed as a means to facilitate and hasten transition from the agricultural to the industrial stage of national economy. A nation on the threshold of this change will

be taking the first steps toward establishing numerous manufacturing industries. In these first steps manufacturing will have to contend with the competition of imported goods coming from older nations whose industries are already firmly established. They have the machinery, the technical knowledge, the skilled workers; they are directed by experienced entrepreneurs. None of these essentials has as yet been attained by the young industries of the first country, and the competition of foreign manufacturers is felt as a severe handicap.

Under these circumstances the tariff may be used to give relief from foreign competition until such time as the young industries have become firmly established. This line of argument is known as "the infant industry argument" for protection. When logically used it rests upon the following assumptions: (1) the argument applies only to industries which are suited to the national conditions; (2) the obstacles to the establishment of such industries are temporary only; (3) the tariff protection need be temporary only, and after the removal of the tariff the industries will be able to stand alone.

When these conditions are fulfilled, and when it is further proved that the nation will actually benefit from the establishment of the industries in question more rapidly than they would come in the natural order of events, the infant industry argument is undoubtedly sound.

Conditions in the United States at the beginning of the nineteenth century were fairly typical of those here assumed, and the first tariff acts were perhaps to some extent justified by this argument. Even more than the tariff, the international conflicts of the period from 1807 to 1815, by interfering with foreign trade, gave protection to the young American industries. But the conditions to which the infant industry argument was applicable had generally passed with the first third of the nineteenth century. This argument is little heard of today.

It must be noted also that the infant industry argument has not often been urged in good faith; having used it to obtain tariff favors, its advocates have later shifted to other arguments to prevent removal of protection. Most protectionists today advocate permanent protection.

Protection against "dumping." From time to time producers of some nation undertake to unload a surplus product upon foreign

markets at prices lower than are charged in their own domestic markets, sometimes even lower than cost of production. When goods are thus exported at less than cost the process is called *dumping*. Dumping may be resorted to by some particular firm whose position sufficiently approaches monopoly, or by a combination of producers acting in concert for this purpose—for example the one-time closely organized chemical industry of Germany. Often such export trade is encouraged or actually subsidized by the government. For a long time it was the practice of the state-owned railroads of Germany to grant to certain great cartels special low freight rates on goods destined for export.

During the decade of the 1930's dumping became more and more a governmental policy, especially of certain authoritarian nations, employed in connection with their efforts to control their foreign trade and foreign exchanges and protect their stocks of monetary gold. Similar results are caused when nations, as has been quite common in recent years, deliberately devalue their currencies in order to take advantage of the temporary stimulation to export trade which tends to follow.

The producers of any nation whose markets are invaded by such dumping of foreign goods are subjected to competition of special severity. Were this a stable and presumably permanent condition, it would be the part of wisdom for the nation concerned to accept the situation, congratulate itself that its people were able to obtain some needed products thus cheaply, and let its own capital and labor transfer to more advantageous lines.

But dumping, from its very nature, is fickle and unstable. Such temporary competition may work serious injury to a nation's industries, only to leave its people poorly supplied when the foreign producers return to a more normal price policy. The reader will recognize the similarity between foreign dumping and certain practices of monopolistically inclined enterprises in the domestic market —practices which have earned the title of "unfair competition."

It is only natural that a nation should seek to protect its own producers against dumping. The protective tariff is the obvious means. Wisely employed, the tariff may serve to offset the price-cutting of foreign producers, with the result that the foreign goods enter the domestic market at prices not less than their normal costs of production. This policy involves frequent adjusting of tariff rates to meet changing policies of foreign producers, and it

requires skillful handling. There is moreover always the danger that selfish interests will use the dumping bogie to secure excessive or permanent protection. Suitably employed, however, protection against temporary dumping is in the interest of the whole people of the nation concerned. Protective tariffs and other measures against dumping are at present being employed by an increasing number of nations.

State controlled trade. A still more serious obstacle to the operation of free trade appears when the people of one nation must trade with the government of another nation. Russia furnishes an example; her government exercises, through the Amtorg Trading Corporation, a monopoly of the nation's foreign trade. The exporter who undertakes to sell in Russia can sell only to Amtorg; the importer of Rusisan goods can buy only from Amtorg.

We have been studying the economic forces which, under freedom of trade, tend to bring equilibrium in the trade between nations, with the maximum of advantage to all concerned. This analysis, however, assumed that trade "between nations" was actually carried on by private persons in the respective nations—not between their governments, nor between private persons in one nation and the government of another nation.

In the latter situation it may well be that the interests of the nation where free private enterprise prevails cannot be safely entrusted to the tender mercies of free trade between its own nationals and another people's totalitarian government. The latter may be able by arbitrary actions—now dumping goods at less than cost, later restricting or prohibiting certain exports; now refusing to accept all or certain foreign goods, later subsidizing their import —to throw the first nation's foreign trade into turmoil and adversely affect the functioning of its domestic economy.

It is conceivable that under such circumstances as these, a nation, while recognizing the normal advantages of free trade, may find it necessary to sacrifice freedom, and, from time to time, exercise some control over its international trade, in order to protect its own interests.

Protection on account of military necessity. Protection may be urged from an entirely different standpoint. The tariff is necessary, it is said, in order that the nation may be assured of a supply of certain products vital to its safety in time of war.

If it be proved (1) that a certain industry is thus necessary and

(2) that this industry will not exist unless relieved of foreign competition, then it follows unquestionably that protection is advantageous. Whether these two assumptions are actually fulfilled is a matter of fact to be determined in each case; it is doubtful whether, in a country having the magnitude and variety of resources and industries possessed by the United States, there is much room for this line of argument.

Without doubt the extraordinary development, in the period between the two World Wars, of protectionism among the nations of Europe was in large measure based upon military considerations. These nations were preparing for war and seeking the utmost of national self-sufficiency in order that in the next war they should not be weakened by enemy blockades, as was Germany, for example, in the First World War. Thus we have seen tariff rates of unprecedented height, embargoes, quotas, and the other devices which marked the nationalistic hysteria of the second quarter of the twentieth century. As a result of these policies the outbreak of war in the autumn of 1939 saw the chief belligerents possessed of a far greater degree of national self-sufficiency than they had enjoyed in 1914, though doubtless at the cost of sacrifice in the living standards of their people.

It should be noted that the military necessity argument for protection is not based upon the traditional economic grounds. It does not rest theoretically upon the false assumption that the people will be richer on account of the tariff. It can admit that they will be poorer, through having to pay more than is necessary for some important commodity. But so are they made poorer by the taxes which defray the cost of building battleships. The justification is the same in either case.

• **Protection by bounties.** Occasionally the object of protection is attained, not by a duty on the foreign product, but by a bounty paid by the government to the domestic producers. The United States, for example, has on occasions sought to encourage the growing of sugar by means of bounties. As to their economic significance, the bounty and the protective tariff are alike, with one important exception. The cost of protection by tariff is imposed upon the consumers in proportion to their use of the goods in question. The cost of the bounty rests upon the whole group of taxpayers. Since protection can be justified, if at all, only on the ground of public advantage, the apportionment of the cost under the system of boun-

ties is more equitable than under that of the protective tariff. Nevertheless bounties are very rarely used, probably because it is here so evident that protection costs something, that a burden is placed on the whole people for the benefit of a favored few.

Ship subsidies. Ship subsidies are bounties paid to ship owners or operators in order to encourage the development of national shipping—the merchant marine. The economic nature of ship subsidies is precisely the same as of protective tariffs and bounties. The justification of subsidies to shipping must be found, if at all, in the military necessity argument.

International trade agreements: Preferential arrangements. At the same time that the recent wave of extreme nationalism was leading most nations to curtail severely their trade with other nations, there developed a contrary tendency to enlarge the area of comparative freedom of trade by means of special tariff agreements.

The British Empire furnishes a conspicuous example. In 1932 Great Britain took definite action terminating her traditional policy of free trade, and in the same year established a scheme of imperial preference by agreement with her various self-governing dominions. Thus was the British Empire bound together in an area of tariff preference as against the rest of the world. France, without going so far toward a formal organization, made preferential tariff arrangements with her colonies. Germany was seeking to build up a preferential tariff group, including nations under her special influence in central Europe as well as certain South American countries. Japan was seeking to dominate a great Eastern tariff area including Manchuria and China.

Reciprocal trade agreements of the United States. A somewhat different line has been followed by the United States. The trade agreement act of 1934 (continued by acts of 1937 and 1948) authorized the President to enter into foreign trade agreements with other nations, in connection with which he was empowered to make reductions up to 50 per cent in existing tariff rates upon certain imports from a given nation in return for corresponding concessions in that nation's tariff in favor of American exports.

The State Department, from the first, acted with vigor in negotiating such reciprocal tariff arrangements. By the fall of 1947, not only had many such binational agreements been made but a multinational agreement involving twenty-three other nations, including Great Britain and Canada, had been concluded at the International

Conference on Trade and Employment held in 1947 at Geneva, Switzerland.

These agreements are reciprocal, but not preferential like those of most other nations, since the United States still lived up to the liberal "most-favored-nation" principle, in accordance with which any reduction granted to one country is "generalized" to all nations with which the United States has most-favored-nation treaties. This does not mean that there is no advantage in being the party to a trade agreement with this country, since the agreements are based on the "chief-supplier" principle; *i.e.*, the United States makes tariff reductions only on articles coming exclusively or chiefly from the nation which is the other party to the agreement.

It would be too much to suppose that this reciprocal agreement policy is making any real breach in the traditional wall of American protectionism. Indeed care has been taken not to permit competition that would seriously affect any American industry. This is accomplished both by the choice of articles for tariff concessions and by limiting the amount of import; for example, in the agreement with Czechoslovakia her producers were permitted to send shoes at the reduced rate, but only up to 1¼ per cent of the average annual production of the United States.

Nevertheless in a world that has been rapidly throwing away important economic benefits through an exaggerated nationalism, the recent attitude of the United States has presented a powerful influence on the more liberal side. This comes with special significance from the nation which for a century has been the chief exponent of the policy of high tariff protection.

War and restriction of foreign trade. The movement toward mercantilist restrictions upon foreign trade, which was gaining momentum in the decades following the First World War, was of course tremendously enhanced by the Second World War. During this conflict military considerations, to a degree never before witnessed, impelled all the nations to seize virtually complete control of their imports and exports, sweeping away nearly all remaining vestiges of freedom. Tariffs, quotas, embargoes, preferential trade arrangements, discriminatory import and export duties, subsidies and penalties, and outright buying and selling by governments, often assuming monopoly position, were the order of the day. Here was the twilight—not to say the total eclipse—of free trade.

Following the war the pendulum began its swing in the other di-

rection, though only slowly and weakly in most parts of the world. America again took the lead, expanding her previously inaugurated policy of reciprocal trade agreements, and seeking, by means of agreements among the nations and through conditions attached to her grants of financial aid (as in the British loan and financial agreement in 1946), to induce other nations to remove or relax the trade restrictions that survived the war.

CHAPTER XXXI

The Business Cycle

Business fluctuations. Character and types. In our modern society economic forces do not remain for long in a state of stable equilibrium. Changes are constantly occurring in some branches of activity that do not quickly harmonize with corresponding changes in other branches. At times the whole economy is expanding and business is generally prosperous; at other times there is contraction and widely diffused recession in business activity. In prolonged periods of depression, as in the 1930's, the lack of adjustment may be so pronounced that unemployment runs into millions, while total production and national income decline disastrously.

Periods of depression create serious social and economic problems. General unemployment means a marked and fairly general drop in the standard of living. The homes and savings of many are lost; others lose work skills and become afflicted with a deep-seated pessimism to which there appears no limit. Economic growth, increasing division of labor, and the accumulation of capital goods are retarded, and technological progress may be sharply inhibited. At such times doubts easily arise as to whether the system of private enterprise can ever provide reasonable stability of employment and production, and apostles of authoritarianism, in one form or another, find it easy to sow the seeds of discontent.

When we speak of an economic fluctuation, we mean a change of business activity on a fairly broad scale, which can be represented in statistical terms. The statistical index chosen to represent the change may be selected from among many that are available. For example, we may chart the volume of agricultural and factory production over a period of time, or the extent of employment of wage earners, or the average price level, or the rate of return on invested funds, thus finding evidence of a general movement upward or downward as the case may be. In the following pages, we shall use such indexes as these to aid us in the analysis of the problem of economic fluctuations.

Economic fluctuations may be classified in four main categories:

(1) secular trend, (2) accidental irregularities, (3) seasonal cycles, and (4) business cycles. Of these types of fluctuation, the fourth—the business cycle—is by far the most important, both as a matter of practical concern to mankind and as a problem of economic theory. This chapter, therefore, is devoted chiefly to the business cycle, with only a brief comment on the other types of fluctuation.

Secular trend. The term "secular trend" refers to the long-run growth or decline of an industry or an economy. So far there is little information, either statistical or theoretical, on the nature of the trend.

Some economists believe that the secular trend is a long-run business cycle. Its upward phase lasts from forty to fifty years, and the downward phase in a progressive economy rarely lasts more than thirty years. It is also believed by some that the severity of the depression of the 1930's is to be explained by the coincidence of the recessive phase of a shorter business cycle with the beginnings of a downward secular movement.

Accidental irregularities. This term is used to designate such disturbances in the normal development of economic enterprise as result from isolated events, such as widespread strikes, national disasters, or popular reactions to changes in governmental policy. These disturbances are usually short-lived and self-correcting, but their effects appear on any long-run chart of business activity as irregularities not to be accounted for by reference to the major forces of change.

A few economists have tried to show that some accidental irregularities may incite a basic cyclical movement, and that either the upward or the downward phase of a business cycle may be temporarily accentuated through contact of basic forces with accidental disturbances. It has been suggested, for example, that the 1928 recovery in American agriculture was initiated largely by the concomitant appearance of a bad harvest in Europe and an exceptionally good harvest in the United States and Canada. The good harvest brought renewed prosperity to the farmer here, and his improved condition was soon reflected in improved general business conditions. Most economists, however, do not attach so much importance to the harvests as affecting general business conditions.

Seasonal cycles. These fluctuations result in part from market response to the changing climatic seasons, in part from the conventions of our calendar. The increased business activity which cor-

responds with seed time and harvest in an agrarian country is an example. So is the business rush during the Christmas season as contrasted with the slack of midsummer.

Definitely cyclical in character, and appearing each year with measurable regularity, seasonal fluctuation may be regarded as a distinct form of the business cycle. Many students of the seasonal cycle, however, believe it is entirely independent in origin of the longer phenomenon, with its alternating phases of prosperity and depression. Nevertheless, given the right phase of the seasonal and cyclical movements, it may be possible for seasonal cycles to act as initiating or reversing forces in the business cycle. To illustrate, a seasonal upswing occurring at the time of an incipient recovery may hasten the cumulative upward movement of the cycle.

Seasonal fluctuations do not present a major social problem. It is true that certain specific industries—such as the manufacture of clothing, coal mining, the building trades—may be seriously affected. In such industries, the problem is one of stabilization through improved business organization and management. In society generally, the existence of seasonal fluctuations is clearly recognized, and economic activity as a whole is fairly well adjusted to them.

The business cycle. Characteristic features. The term "business cycle" is ordinarily used to describe a complete cycle in business activity, lasting over a period of years. A complete business cycle normally embraces within its compass four well-defined phases, to which it is customary to give the names: *revival* or *recovery, prosperity, recession,* and *depression.*

With minor differences in detail, a general pattern is seen to apply to all business cycles. The revival is usually taken as the starting point and is marked by the upturn of indexes of business activity. The prosperity phase is defined by the later stages of this upward trend. Recession begins when the curve of business activity starts downward. Depression is indicated by the subsequent low points on the curve, continuing until another revival sets in.

The conditions prevailing in the successive phases of a cycle may be roughly summarized. In revival we find rising prices of securities and commodities, a rising volume and value of industrial production, increasing sales and forward orders in most lines of trade, increasing employment, mounting loans or debts, and a rising rate of interest.

In prosperity we find prices continuing to rise but gradually leveling off, large accumulated stocks of goods, high but stabilized employment, an intensified demand for credit accompanied by high rates of interest, and the production of capital goods beginning to weaken.

In recession, the business picture is one of decline in sales, orders, production, employment, and prices, an increasing real burden of debt as the amount of debt is reduced more slowly than the decline in the value of assets, and a falling rate of interest.

In the depression phase, debts and stocks of goods have been reduced; prices run at low levels, though security prices may show some slight return of strength; the rate of production is low; unemployment is widespread; interest rates are low.

Any one of the major indexes of business activity will show the existence of these more or less rhythmic recurrent fluctuations.

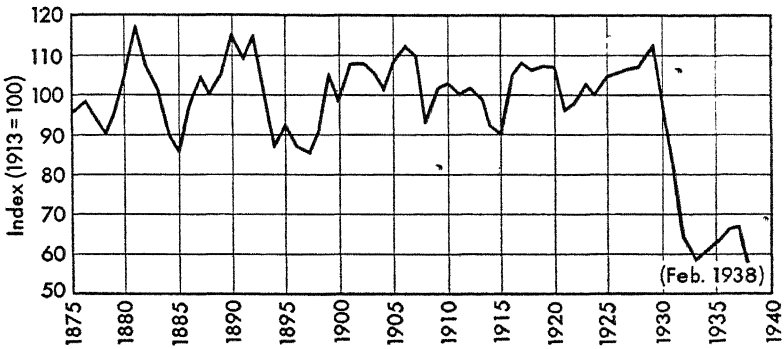


FIG. 35. INDEX OF GENERAL BUSINESS ACTIVITY, UNITED STATES, 1875–1938 [1]

Figure 35 portrays the movements of a general index of business activity over the period from 1875 to 1938. Starting from each low point, there is an upsurge in activity until the peak of prosperity is reached, and then a period of contraction until a new low point is determined. Despite many obvious dissimilarities, it will be seen that the trend of business activity as a whole does describe a rather definite pattern within certain well-marked periods. Measuring these periods by successive low points, we have the following series

[1] Clearings index of the Federal Reserve Bank of New York, with adjustment for secular trend, seasonal variation, and changes in the general level of prices.

of business cycles: 1878–1885; 1885–1897; 1897–1908; 1908–1915; 1915–1921; 1921–1933.

In Figure 36 the movements of the indexes of factory employment and factory pay rolls are shown. This chart covers a shorter span of years than the one preceding, and it is therefore possible to plot the changes, month by month, within the course of each year. An examination of this chart will show the existence of complete cyclical movements of a shorter duration than can be seen in Figure 35. It will be noted that the indexes rose from a low point in 1919 to a high point in 1920, and then dropped to a new low point in 1921. Beginning there, the process repeats itself, another high point being reached in 1923 and another low point in 1924. Then comes a slow rise to a peak in 1929 and an abrupt fall to a new low point in 1933.

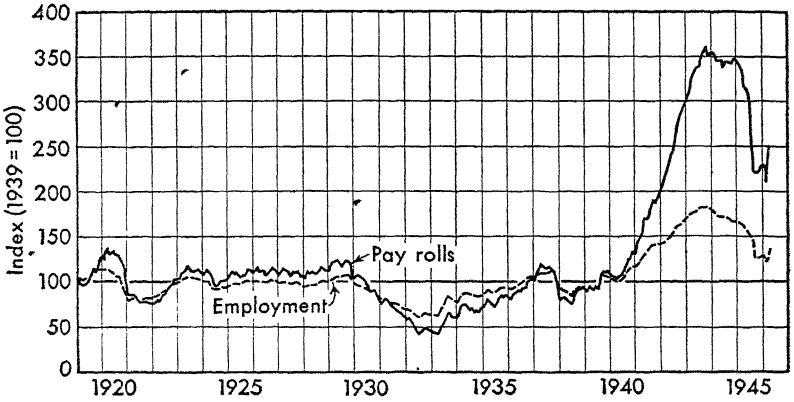


FIG. 36. INDEXES OF FACTORY EMPLOYMENT AND FACTORY PAYROLLS, 1919–1946. (Monthly average 1923–1925 = 100) [1]

This is a record of a series of what are sometimes called minor cycles, to distinguish them from the major cycles already described, which require a somewhat longer period of time before the cyclical forces have spent themselves. In this case it will be noted that the duration of these minor cycles is as follows: 1919–1921; 1921–1924; 1924–1933. Another glance at the chart will show the longer rhythmical swing, starting with an extremely low point in 1921 and

[1] Indexes are published currently in United States Bureau of Labor Statistics, *Employment and Payrolls*, and *Monthly Labor Review*. Washington: Government Printing Office.

ending with another very low point in 1933. In the discussion that follows, we shall pay attention to the longer cycle as probably the more important, but it should be realized that business cycle analysis may be applied to explain either the shorter or the longer variety.

Varieties of business cycles. While the paths that business activities trace through successive cycles give evidence of considerable similarities, at least in rough outline, it must not be inferred that one business cycle follows with any precision the pattern set by the preceding cycle. Business cycles cannot be defined as perfectly symmetrical movements, expected to reappear again and again with the same shape and timing.

Indeed when we picture the business cycle in statistical terms by plotting a number of indexes of business activity—volume of production, employment, national income, consumption, savings, investment, prices, interest rates, wages, and stores and inventories of goods—careful comparison will show that the patterns so described [1] are much more conspicuous for their variations in form than for perfect uniformity. The indexes differ from each other within the same cycle, and a comparison of the same index in successive cycles will likewise show lack of uniformity.

The most obvious difference between cycles lies in their duration. Comparison of the period 1878–85 with any of the other periods in Figure 35, or with the shorter cycles portrayed in Figure 36, will make it evident that no given period of years can be designated as the normal duration of a business cycle, whether major or minor.

Neither are any two cycles precisely alike in the amplitude of fluctuations; in one cycle business as a whole may progress at a faster rate and reach a higher peak than in the preceding cycle, or the downward movement of business may become more rapid and a lower level of activity be reached.

The ups and downs in the indexes of business activity vary from industry to industry in different cycles. In the 1920's, activity in shipbuilding, anthracite coal mining, and railroad transportation lagged behind the majority of industries; while radio, automobile, and electric power production advanced very rapidly, setting new records of upward change in industrial activity. If there is a building boom during an upward phase of a business cycle, residential

[1] With the seasonal and trend effects removed by appropriate statistical techniques.

construction has often led general business expansion. On the other hand (as during the 1930's), there may be very little residential construction, while other types of construction are leading the recovery.

In some cycles the index of employment may decline more than the index of physical production. In others the total national income may decline less than some component of national income—say farm income or business profits. Retail prices may remain more stable than wholesale prices, either in the expansion or the contraction phase. Another cycle may be characterized by a long lag of wages behind the prices of goods and services. In the decade of the 1930's, wages in a number of trades and skills did not fall on the average as much as formerly, and within a year or two from the depth of the depression, they rose more rapidly than other wages and than most prices of goods and services. But in the same decade, wages of unorganized workers did not lag behind prices, but fell even more precipitously.

Analysis of the business cycle: General. This variability of all the measures of business activity, both within each cycle and as between any two cycles, is so pronounced that some have questioned whether there really is such a thing as a business cycle. Certainly the answer would be in the negative if it were found that the explanation of the chain of events in one cycle had no applicability to the situation found in other cycles. But while it is true that in a certain sense each business cycle is unique, it is also true that analysis of the causal relationships within any given cycle is, with such modifications as are dictated by variations in the economic setting, generally valid for other cycles.

The successive phases of the business cycle (except revival, which is usually taken as the starting point) seem to be self-generating. The revival phase develops through its own momentum into the prosperity phase; the prosperity phase contains within itself forces which initiate recession; recession in turn develops through its own characteristics into depression. In our exploration of the causal relationships within a business cycle, it is convenient to start with the period of revival.

Revival. The most significant feature of the revival is an increase in spending. Ordinarily we find that business spending increases first and is followed subsequently by an enlargement of consumer spending. Economic science has as yet no completely satisfactory

answer as to what the forces are which induce this increase in spending at a time when business is depressed and unemployment is widespread. There are, however, some points on which there is general agreement.

In the first place, depression itself to some extent serves to prepare the way for a revival. Weak firms and weak banks, which have survived the first months of the depression, will eventually be forced out through bankruptcy and dissolution or reorganization; and the surviving firms, with assets written down and capital structures reorganized, will be better credit risks and so in better position to secure funds for spending whenever business spending appears to be warranted.

Furthermore the continued deflationary pressure of the depression itself is almost sure to bring about a reduction in costs. During a depression, wage rates, interest rates, and many of the prices of producers' goods will have been reduced. Not only are costs, both of current production and of new investment, lowered in this fashion, but the less efficient members of any working force and the less efficient producers in any given industry are likely to be the ones first thrown out of employment.

The favorable effect of these reductions in cost should not of course be exaggerated. For one thing, it is quite clear that most reductions in costs bring about equivalent reductions in the incomes of other members of the economic system, so that unfavorable effects on purchasing power may tend to counteract the favorable effects of cost reductions.

Nevertheless if the depression has the effect of decreasing certain sorts of costs that have been out of line with the general level of prices and incomes of the community, it may make the prospects of future profits more favorable and thus encourage increased business spending.

But more important is the fact that, if business is to maintain even a low rate of production, the time will come when it will require new machinery and equipment to replace that which has worn out. During a period of depression it is the practice of business firms to reduce expenditures for maintenance of plant to the minimum by postponing replacements of worn-out equipment, but this process of using up capital goods cannot continue indefinitely; at some time increased spending for this purpose will be inevitable.

All three of the causes of revival so far discussed are in some

degree automatic in their operation, and they have considerable significance. But every history of the business cycle makes it clear that the most powerful forces working to initiate revival are ones which do not develop automatically out of the preceding phase of the business cycle, but are in a sense fortuitous.

For the moment, let us assume that business activity has been stimulated by augmented business spending, and let us follow through the developments characteristic of the successive phases, reserving for later consideration the factors that may have provided the impetus for this.

Prosperity. Once a revival is under way, it tends to become cumulative, feeding on itself for a considerable period at least. It is not difficult to see why this is so.

As new construction projects are launched and carried through, men will be employed in building factories, installing machinery, and other such operations. Part of the increase in expenditure will thus be matched by increases in wage payments. Many of the new expenditures will go toward the purchase of building materials, machinery, tools, etc., and the business firms which supply capital equipment will thus find the demand for their goods increased and their incomes enlarged. They in turn will find it necessary to place orders for raw materials, to engage larger laboring forces, and to revise their production schedules upward.

Consumers who have benefited from the primary increases in income will of course increase their expenditures for consumers' goods, and the demand for these goods will be swelled still further as the larger consumer incomes generated by renewed activity in the capital goods industries manifest themselves. As wage earners come to feel that their jobs are secure and their incomes likely to increase, they incline to expand their expenditures beyond their immediate means and so resort to borrowing in one form or another.

This factor is of great importance for those industries which produce durable consumers' goods, such as automobiles, refrigerators, washing machines, and the like, and especially for the field of housing. During the decade of 1920–29, consumer spending for the purchase of new houses was of very large dimensions and was largely financed by borrowing from the banks.

This substantial increase in the demand for consumers' goods of course gives rise to increased production and employment in the

consumer goods industries and still further reinforces the trend toward increased business activity.

Furthermore investment in stocks, both of raw materials and of finished goods, is certain to be stimulated. During a depression, businessmen almost invariably attempt to reduce their inventories to the lowest possible level, but once revival is under way, this entire picture changes. The expectation that the volume of sales will continue to rise and the fear that a shortage may develop and that deliveries of goods may be slow will induce businessmen all along the line to build up larger inventories than they have been carrying.

Rising prices of goods of all sorts give them an added incentive, for even though they may be most anxious to avoid speculation in inventories, the expectation of a continued rise in prices will make it seem prudent to add substantially to the volume of goods carried in stock. The initial increase in the demand for both producers' and consumers' goods tends thus to be magnified.

As recovery proceeds, and especially if it proceeds rapidly, the increasing demand for the use of the factors of production and for both consumers' and producers' goods will bring about higher prices. But it is characteristic of a period in which the general level of prices is rising that many of the prices which determine costs of production—such as wages of labor and prices of capital goods—lag behind the sale prices which determine business income. The result is a wider margin of profit, which serves as an added stimulus to expansion and employment and therefore to income.

It will be recalled that, when revival starts, banks have large idle cash reserves and are ready to lend at low rates of interest. The volume of investment in the instruments of production is therefore not limited by the quantity of savings available from current income, but can be considerably in excess of that amount. New purchasing power is injected into the system by the expansion of bank deposits as businessmen increase their borrowing, and the more optimistic atmosphere encourages consumers to bring some actual cash out of hoarding. This increase in investment means, in a succeeding period, increased incomes in the way of wages, interest, rent, and business profits, and again there can be an increase in total spending, and therefore a further impetus to business activity.

In summary, we may say that the initial increase in spending gives rise to a primary increase in income. But this in turn stimu-

lates an increase in production over wide areas of the economic system, and the increase in production in turn gives rise to a secondary increase in income. Even there, of course, the process does not stop, for the enlarged output of capital goods gives rise to an increased demand for the raw materials which they require, and hence to further increases in income. Thus the initial increase in spending spreads throughout the economic system, and with its spread the level of production rises, consumers' and business incomes are increased, further spending follows, and production is again stimulated.

Recession. Prosperity culminates in recession by a corresponding set of causal relations. The interaction of increased incomes and spending and expanding business activity, which characterizes the prosperity phase, may continue as long as income recipients either spend or invest their entire incomes, and as long as banks continue to expand deposit currency. When conditions develop to arrest or stop either of these two processes, the interaction necessary for prosperity is interrupted, and the first step toward recession takes place.

As prosperity proceeds and consumer incomes rise, the rate of consumer spending may decline, for as incomes increase the natural tendency of many income recipients is to save a larger portion of their enlarged incomes. As long as these increased savings are invested in their entirety in new plant and equipment, they are disbursed to those employed in this capital formation and therefore do not necessarily reduce the rate of total consumer spending. But if, for any reason, the opportunities for profitable investment shrink and savings result in a contraction of bank credit and spending, the sustaining factor of prosperity weakens.

Continued expansion of deposit currency is conditioned, first, upon the confidence of entrepreneurs that loans can be used profitably, and secondly, upon the ability and willingness of the banks to expand their loans. Eventually legal requirements limit the ability of banks to expand deposits, but before these restrictions become operative, the rise in the interest rate and the more careful scrutiny of the purposes for which loans are requested retard the expansion or bring it to a halt.

As the investment opportunities of the prosperity phase are exploited, the prospect of further profitable investments declines. Sooner or later the rise in prices of industrial products will slacken,

while wages, interest, and other more sluggishly moving prices which affect costs catch up. So it is that the margin of profit on sales is reduced, and entrepreneurs are then called upon to determine whether future profit possibilities warrant further plant expansion.

Under these conditions, businessmen, including bankers, may conclude as the first possibility that industry has already overexpanded and may drastically curtail investment in capital goods. Then consumer savings will not all be absorbed by investment, bank loans will shrink, and a recession in business will be ushered in.

Second, it is possible that businessmen may decide merely to reduce the rate of plant expansion, in the belief that, though there may be room for it, at least in some lines, a more leisurely pace might bring about better coördination between the various parts of the economy. If this were to be done in such a way that total spending was not reduced, there would be no reason why consumers' savings could not continue to be invested, nor why the volume of bank loans should be reduced.

Finally, the verdict may be that the decrease in the margin of profits is but a pause in the march to a future of boundless promise. If there is this degree of optimism, investment may continue far in excess of savings from current income, so long as credit facilities are available. If this situation continues long, the final stage may well be overexpansion of plants beyond the possibility of profitable operation in the immediate future, speculative bidding up of the prices of securities on the stock exchange—to a point above any reasonable level based on discount of their future earning power, and speculative booms in farm lands (Florida real estate and the like), until a point is reached where collapse is inevitable.

The question may well be raised as to whether it is not possible—when the point of full production has been approximated and before competition for raw materials and workers has forced up their prices to unprofitable levels—for business activity to level off and to be sustained with reasonable stability on a new and higher plateau. Though there are many who believe that this can be achieved by a careful and intelligent rationing of credit, the fact remains that to date we have not succeeded in doing this. In past periods of prosperity, excesses of one sort or another have occurred, which seem to have put the whole economy out of joint, with disequilibrium in the price structure and in the structure of industry itself.

Depression. At the culmination of a period of prosperity, the developments we have traced finally interact so as to bring a reduction in the rate of prospective profit. The profit margin is pinched between a retarded rise of selling prices in general and an accelerated rise of costs. As the rate of interest finally rises above the rate of prospective profit, the incentive to borrow for expansion is replaced by an incentive to curtail business spending and pay off loans. As soon as the industries producing capital goods find a definitely declining market, they add an absolute reduction of output, disbursements, and bank loans to the now rapidly changing picture of business. With consumer spending then further curtailed, a definite decline in industrial activity becomes general, and profit margins are further narrowed as business costs tend to fall less rapidly than other prices.

Thus begins a spiral of falling values, forced sale of goods to permit required debt liquidation, consequent reduction of credit currency, further depression of values, increased unemployment, and a general decline in incomes and trade. A psychological depression results, and a reduced velocity in the turnover of currency augments the other factors. Declining prices of securities on the exchanges add further to business curtailments, partly by their psychological influence, partly because of their effect on the credit standing of business enterprises.

As trade approaches a minimum and surplus stocks of goods are exhausted, as the burden of old debts is slowly removed, as the rate of interest falls, as costs of production fall into line with the general decline of prices, profit margins tend to increase enough to permit the use of funds more profitably than in paying off loans. Such new funds might be used for slight expansions here and there in the volume of operations, or for meeting the requirements of postponed maintenance of capital equipment. A period of uneven but general equilibrium between costs and selling prices, between production and consumption, thus tends to prevail until events external to the conditions of depression occur to stimulate another revival.

Forces initiating the business cycle. Apparently the conditions internal to the business situation in a period of depression can at most set the stage for a possible revival, leaving it to external forces to set the cycle in motion. The normal operation of the credit system exerts no original impulse upon the cycle; low interest rates,

excess bank reserves, and superabundant supplies of credit proved ineffective in 1932 and 1933 to bring revival out of depression. Similarly the equilibrium of costs and income, which obtains throughout the business world after recession has worked out its effects, is only an invitation to accelerated activity. The spur of self-interest under such conditions certainly moves entrepreneurs constantly to seek opportunities to expand their businesses. These impulses may result in a degree of growth on a somewhat uniform level, but they do not appear to be effective for starting the upward accelerated surge of activity which is characteristic of a cycle in its first phase.

As a rule, depressions have not ended until substantial new investment outlets have been discovered. Historically the classic example during the nineteenth century was the development of railroads throughout the world. In the United States, up to 1890 at least, this provided the largest single outlet for investment funds. The existence of such investment outlets affected not only the industries directly and indirectly concerned with the railroads themselves; the high rate of investment in railroads made for a high national income and thus stimulated investment in many other lines.

A more recent example is the automobile industry in the United States, which during the decade of 1920–29 grew at a prodigious rate. Founded many years previously, the industry had evolved by 1921 to the point where it presented an ideal opportunity for the investment of capital on a large scale. This investment opportunity and the shortage of housing that had developed during the war were largely responsible for the quick ending of the 1921 depression.

Investment opportunities may take several forms. Though perhaps the most important ones occur when a new industry or a new area begins to develop, much investment is undertaken by existing firms in existing industries and without reference to the opening up of new territories anywhere in the world. Investment frequently occurs as the result of new and cheaper methods of production, such as are being discovered almost continuously. After a depression has been in progress for several years, even if no major new investment opportunity appears, replacement of old equipment and expansion of plant are eventually forced upon businessmen, both by the wearing out of what they already had, and by the availability of new kinds of machinery and new productive methods.

Theories of the business cycle: The depression of the 1930's. Until the 1930's, there was little disposition to doubt that sooner or later the blight of a depression would be lifted by the impact of some external force, but the depression which started with the crash of 1929 did not run true to form. Depression deepened for a number of years, reaching its nadir in 1932, and although enormous government expenditures were made in the hope of starting a revival, such modest recovery as appeared was a very fragile thing, as shown by the relapse in 1938 following the curtailment of government spending. Full employment did not come until the demands of the military services and war production plants drained the pools of unemployed and created instead a labor shortage.

There is general agreement that the basic difficulty was that, for a decade, profit expectations were not sufficiently alluring to induce private investment on a scale broad enough to support a vigorous recovery. The evidence as to the drop in capital formation is quite clear; in 1929 gross capital formation reached a peak of 20.7 billion dollars, but by 1932 it had fallen to 3.8 billion dollars, a decrease of nearly 17 billions. While there was a recovery in subsequent years, it was slow, and even in 1937 the figure was but 71 per cent of that in 1929. (See Figure 37.) Business cycle theorists are still not in agreement as to why, when the bottom of depression had once been reached, revival did not appear until forced by the demands of war.

Theory of the matured economy: Statement. One school of thought professes to believe that our economy has reached the stage where it is futile to expect external forces to lift business activity from the depths of depression, or even to expect internal forces to carry revival on into prosperity. In support of their theory, the followers of this school make two major assertions.

The first is that in present-day America we have a "matured" economy. The growth in population has slowed down; the frontier has come to an end; there is a dearth of technological improvements; and in general there is evident a tendency for economic and other forces to lose the dynamic qualities characteristic of them in the America of yesteryear. The result of all this is said to be a marked decrease in investment opportunities.

The seriousness of such a condition, if it were true, must be patent. No longer could reliance be placed on the opening up of

new investment opportunities to stir business from its lethargy, or on the cumulative effects of business spending, enlarged by taking advantage of these opportunities, to bring about prosperity.

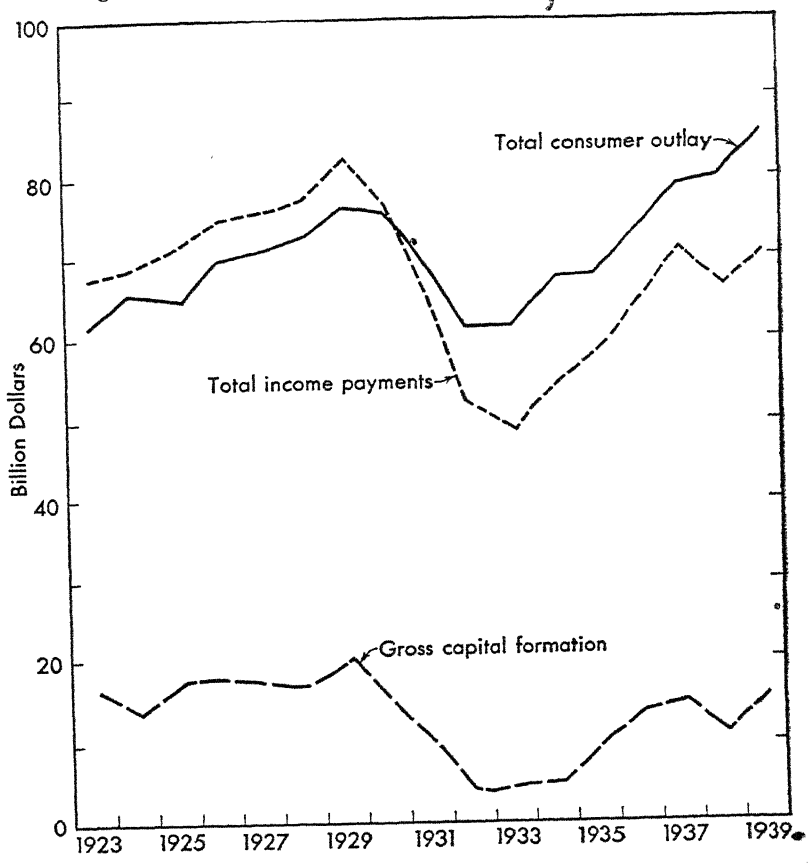


FIG. 37. EXPENDITURES FOR CONSUMPTION AND CAPITAL FORMATION IN THE UNITED STATES, 1923–1939[1]

In addition it is claimed that there is discernible an increasing propensity to save. As individual incomes increase, there probably is a natural tendency to save a larger proportion of the income and to spend a smaller proportion, but the statement is intended also to

[1] Adapted from figures in Kuznets, S., *National Income and Its Composition*, and *National Product since 1869*. New York: National Bureau of Economic Research, 1941 and 1946, respectively.

mean that there has been a noticeable tendency for savings to increase within a given income class; *i.e.*, those earning $5,000 a year, say, are saving more percentage-wise than those who earned $5,000 a year a decade ago, income taxes presumably aside.

A decline in investment opportunities, together with a tendency to increased saving, would presumably mean a decline in total spending and so threaten stagnation, unless spending or investment by others were to make up the difference. Government, so it is said, should therefore be called upon to step in where private industry fears to tread; public expenditures are prescribed for this condition of economic anemia—expenditures in such volume that total spending will be sufficient to support full employment.

According to this theory, if funds for this purpose can be secured by taxation that gathers in the funds that would otherwise be saved and hoarded, so much the better. But if spending has to be financed by budgetary deficits, there is, according to this theory, no real reason for alarm. We are told that these deficits may be continued interminably without endangering the economic life of the community so long as there is a rising national income.

Criticism. One cannot hope to do more than touch the fringes of this controversial question, but it should at least be observed that this analysis hardly seems to fit the cyclical movements of the 1930's. The evidence that the decline in the growth of population and the end of the frontier have been major factors contributing to a decline in investment opportunities is not convincing.

The assertion that continued depression in the 1930's is traceable in large measure to lack of development of new technologies is yet to be proved. The conditions in the 1930's were so complex in nature that there is little data which throws proper light on what did happen in the discovery and use of new technologies.

This cry of alarm that there is a dearth of technological discoveries has been heard intermittently every fifty years or so; it was heard in the 1830's, in the 1880's, and again in the 1930's. Yet even while the cry is being voiced, great discoveries are in the making. The 1830's marked a beginning of the large-scale application of the steam engine to transportation by land and by sea; in the 1880's, there came the application of electricity and the gas engine to many new ways of manufacturing, traveling, and living; and now, plastics, the airplane, the rocket, and atomic energy.

The assertion that there is a growing propensity to save does not

rest on conclusive statistical evidence. On the contrary, the available statistics may point to the opposite possibility; i.e. that relative to total expenditures, consumer expenditures have been increasing rather than declining in the United States and the United Kingdom.[1]

On strictly logical grounds, this observed tendency of a rising propensity to consume may be reasonably supported. We know that, as the economy grows and more advanced techniques are employed to bring prices down through savings in cost, individuals easily acquire desires for goods other than mere essentials. The automobile, the radio, air conditioning, and so on, open up such vistas of good living that the individual wants to go beyond expenditures on necessities. In time what was viewed as a luxury becomes a necessity. This tendency, moreover, has been fostered by the convenience of consumer credit in its various forms—especially installment sales credit —so that millions of individuals may realize their desires by spending in anticipation of income.

A free economy: Recent restrictions. Application of basic economic principles to the events of the 1930's indicates a quite different explanation of the long-continued depression. This is summarized in the remaining pages of this chapter.

Business recovery and business progress depend on a free economy. In the economy of the 1920's, there were many features of monopoly or monopolistic practices that impeded the ready operation of the pricing system.

Moreover many of the attempts in the period of depression to bring about fuller employment—farm price maintenance, the NRA, the sanctions given to labor in the establishment of restrictionist practices, and the like—merely served to accentuate the earlier inflexibilities and to make the system less free and even more unworkable.

Heavy taxes on individual and corporate incomes, defended as a means of promoting consumption by taking from those who would not spend, made the return on investment of capital in new and risky ventures seem frequently too meager to warrant assumption of the risk involved. Thus the incentive to investment was checked.

Furthermore the transfer of spending power to those with the lowest incomes—even granting that in some cases it may have been

[1] Clark, Colin; *The Conditions of Economic Progress.* New York: The Macmillan Company, 1940.

necessary to avoid acute distress—encouraged the consumption of perishable consumers' goods, whereas the essential need was to stimulate industries making durable consumers' and producers' goods.

The free enterprise economy had long been accustomed to capital accumulation on a great scale. Almost the entire institutional structure and entrepreneurial thinking for nearly two centuries had been attuned to a tempo of economic progress involving great capital accumulations. It would seem that the attempt to reverse the habits of saving, production, marketing, and consuming could not be successful. Large industries would have to be changed over to the production of less capitalistic goods, while a general reconstruction of the financial and legal framework might be even more essential.

In all major respects the economy of the 1930's had neither the flexibility nor the temperament for sudden and radical change, and the program contemplated and partly put into operation was probably defeated by amplification of the very forces it was supposedly designed to overcome. The uncertainties introduced by these efforts to change simply compounded the uncertainties growing from a shackled price system. The one choice left to many was to spend largely on necessities; others simply increased their liquid assets. Purchasing power did not flow from such actions in the amount necessary to maintain full employment.

Pump priming and deficit spending. The suggestions contained in the paragraph above deserve some amplification, especially in view of the practical importance assumed by efforts to increase consumer spending in the period of depression. In the early days of the New Deal, it was anticipated that, after the pump had been thoroughly primed through government expenditures, it would work without further attention. In other words government spending was viewed as the external factor—though an artificial one, to be sure—necessary to give the impetus to business revival. It was only when business activity showed little disposition to stay revived that the doctrine of continuous government support, to offset the alleged decreasing propensity to spend, was advanced.

If it be assumed that very little plant is used at capacity at the time such a policy is put into effect, then an increased demand for some or all goods would presumably tend toward fuller use of plant and equipment and make for more employment in some or all of

the plants. The extent to which the increase in purchasing power would be effective would depend, among other things, upon the extent of excess capacity in each industry and the way in which the increases in demand were distributed. Obviously some industries would be closer to capacity than others; if the increases in demand were in proportion to these excess capacities, the probability that the economy would reach a general state of equilibrium might be plausibly argued.

This latter assumption is the one usually (though only tacitly) made when it is held that an injection of additional purchasing power, chiefly placed in the hands of those with lowest incomes, would reduce unemployment and bring prosperity. But these assumptions may not be very realistic. It may be found that the new demands for goods and services are out of line with the excess capacities in the various industries, and moreover that the flow of purchasing power may be in such directions as to create more hoarding than spending. The fact that by 1939 there were still millions of unemployed gives support to this hypothesis.

Restoration of investment. By 1937, following four years of deficit spending by the federal government and of relief payments and public works expenditures by state and local governments, there was a large increase in certain areas of consumer expenditure over the early years of depression. As can be seen in the table on page 634 and in Figure 38 on page 635, the index of total consumer outlay for that year reached 103 as compared with 80 for 1932, and by 1939 the index number had climbed to 111.

With a rise in the consumption of nondurable consumer goods well above the 1929 figure, we might have expected a somewhat comparable advance in the output of both producers' and consumers' durable goods. Actually the index number for producers' durable goods, after sinking to 33 in 1932, rose only to 88 in 1937 and thereafter fell back to 81 in 1939. For durable consumer goods the index dropped to 49 in 1932, rose to 83 in 1937, and then slipped back to 80 in 1939. Total gross capital formation in 1937 was 29 points below that of 1929, and in 1939 it was still 27 points short of the 1929 figure. By 1939 the index for producers' durable goods reached 81, while that for private nonresidential construction—an important indicator of activity in factory, office, and store construction—had crawled up to 33, but two points above the 1929 low. Inasmuch as durable consumer goods require more producer

THE AMERICAN ECONOMY IN TRANSITION, 1929–1939[1]

(All figures are relatives on a 1929 base)

Time Series	*1929*	*1932*	*1937*	*1939*
Gross Natural Product				
Total	100	67	97	103
Consumer outlay				
Total	100	80	103	111
Perishable goods	100	92	122	129
Semidurable	100	81	84	95
Durable goods	100	49	83	80
Services	100	78	99	111
Gross capital formation				
Total	100	18	71	73
Producer durable goods	100	33	88	81
Residential construction	100	27	51	76
Private nonresidential construction	100	31	40	33
Public construction	100	92	88	100
Labor Force				
Total	100	104	110	112
Number employed				
Total	100	81	98	97
Civil nonagricultural	100	74	97	95
Number unemployed	100	794	485	590
Output, Employment, Earnings in Major Industries				
Agriculture:				
Output	100	100	106	111
Number employed	100	96	91	89
Average hourly earnings		(no data available)		
Coal mining:				
Output	100	60	80	73
Number employed	100	81	90	82
Average hourly earnings	100	81	125	129
Manufacturing:				
Output	100	54	103	103
Number employed	100	62	102	94
Average hourly earnings	100	85	118	122
Steam railroads:				
Output	100	52	81	75
Number employed	100	62	68	60
Average hourly earnings	100	95	109	115
Electric light and power:				
Output	100	93	136	152
Number employed	100	84	96	93
Average hourly earnings	100	104	126	184

[1] Source: National Bureau of Economic Research.

goods for their manufacture than do perishable consumer goods, it is perhaps not surprising to find that the index of durable consumers' goods was in 1939 but 80 per cent of the 1929 figure.

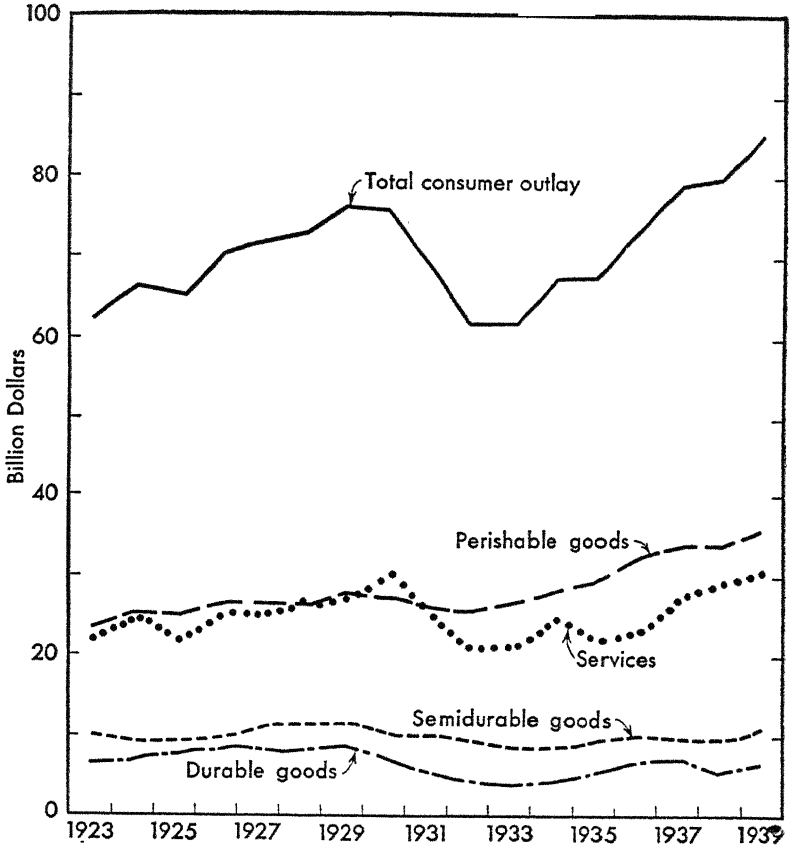


FIG. 38. CONSUMER OUTLAY IN THE UNITED STATES, BY KINDS OF GOODS, 1923–39 [1]

Explanation of this situation may be found in the growing inflexibilities of prices and wages. As we have noted, general equilibrium requires that the supplies of goods and services advance or decline fairly evenly with the increases or decreases in the demand for goods and services. If an advance is impossible in certain areas be-

[1] *Ibid.*

cause of these inflexibilities, decrease in total spending is not necessarily avoided by an immediate and automatic transfer of demand to other lines. Indeed it is more likely that such a situation will encourage a decline in the propensity to spend—the very thing government spending was intended to combat.

An excellent illustration of this is afforded by the history of residential construction during this period. During the last half of the 1930's, the rate of increase of new families rose, while throughout the decade population continued to grow. Yet there was far from a proportionate increase in residential construction. In 1937 such construction was only half of what it had been in 1929, and by 1939, when the marriage rate was rising, the index of the value of residential construction was but 76.

It is not difficult to account for this lag in home construction. Instead of going down—as had the costs of food, clothing, radios, and other consumers' goods—dwelling costs advanced rapidly, in the face of a precipitous decline in rents. To some extent this may be attributed to the heavy concentration of government spending in the construction field, but in large measure it was the result of restrictive practices; *i.e.,* the unwillingness of various building trade unions and contractors to permit the introduction of cost-saving methods of construction, "feather-bedding," and trade union practices which by one device or another forced up wages and other unit costs much more rapidly than in other industries. Thus the rise of wages in the building trades in early 1937 was so rapid as to be unprecedented, and price advances were almost as speedy.

It was stated previously that when prices and incomes in some industries move more and more out of line with prices and incomes in other industries, a compensating increase of expenditures upon the products of industries whose prices lag does not necessarily materialize. There may be two related movements which reduce such compensatory spending by individuals.

The first stems from the inelasticities of the supplies of the factors of production. For example, in the building trades wage increases, offered to bring forth more labor, may be but slightly effective because of the restrictions upon entry into labor unions. The net result may be an increase in wages for those employed and a higher price of the product.

Secondly, the unemployed, already uneasy, are made more uncertain about their future incomes and incline to continue their ex-

penditures at a bare minimum. In other areas of the economy, many producers become uncertain about the future prices of goods and services affected by the restrictive practices and wonder whether prices that seem so out of line can be maintained in the future. Therefore, rather than spend, they prefer to await further developments and in the meantime to increase cash balances.

In the 1930's many individuals reduced their relative expenditures upon durable goods—notably housing, automobiles, and household goods. Others turned to purchases of low-priced and inferior qualities of wearing apparel and household furnishings, and to second-hand cars instead of new ones. Automobile production declined to a point where it barely provided for replacements.

Entrepreneurs, instead of making investments in producers' durable goods, shifted as much as possible to short-term capital, or made only seriously needed replacements of long-term capital.

Durable goods industries in general were probably affected by the rapid advances in the price of steel, perhaps the most important commodity from the standpoint of over-all business activity. The index of steel prices, down to 82 in 1932, rose to 114 by the middle of 1937. This increase probably reflects the rise in steel wages along with unwillingness to pass on, in the form of price reductions, cost-savings resulting from the application of certain improved methods of steel production—particularly new large-scale methods in rolling and shaping sheets and forms.[1]

In other industries—bituminous coal, for example—while prices did not advance to the same extent, they nevertheless did not fall as much as they would have if the savings in cost had been passed on to the entire economy and not devoted to wages and profits alone.

In a word, because of restrictive practices of producers, there may appear in the economy a tendency to delay spending. Investors will be led to wait until a time when the uncertainty about prices and incomes has sufficiently abated to warrant assuming the risks of those obligations which are usually incurred in financing the production and distribution of all durable goods. It is only in this respect that the declining propensity to consume comes to have any bearing upon the cyclical aspects of the price system.

[1] Bell, Spurgeon, *Productivity, Wages, and National Income*. Washington: The Brookings Institution, 1940.

Stabilizing the economy. In a free market economy, with monopolistic and restrictive practices held firmly in check, the price system would probably readily adjust the flow of production to demand as expressed by willingness to buy at given prices. The clearing of the market under these conditions is automatic. The return on investment and the payment for the services of labor would tend to be determined by the markets for these factors rather than by monopolistic control of their respective supplies. There would be a minimum of income inequalities arising out of monopolistic control over supply.

The price system would, moreover, bring adjustment in such a relatively short time that the advance of some industries over others would tend to be considerably less disproportionate and unbalancing. The accidental events of new products, new demands, new technologies, while they would still bring about those changes in the economy which are the marks of progress, would not tend to lead to major disequilibrium. The swings of the economic pendulum would be reduced to short movements about a point of equilibrium, with shifts in demand operating through price changes to bring capital and labor into production of the products that the people demanded.

Curtailment of monopoly and market restrictions may require considerable reshaping of individual behavior, and many persons, looking only at things near at hand, may perhaps respond reluctantly to efforts to curb their practices. Nevertheless, if man is to continue to make his own decisions about the way he works, spends, and invests, he must realize that a program to eliminate monopolistic and restrictionist practices is in the end the true road to prosperity, both for him and for his neighbors. To that end he must focus his attention, along with that of his neighbors, upon the construction of an adequate antimonopoly and antirestriction program, to be administered by a strictly impartial government that is itself limited to its essential governmental functions. This is the only solution which the economist can honestly offer to those who desire to retain the free enterprise economy.

It is clearly an essential part of this program that the government itself refrain from industrial activities which compete and interfere with private enterprise, and likewise that it limit its interference with the freedom of private industry to the restrictions necessitated by its police function and its duty to restrain or regulate monopoly.

Freedom of capital to enter industry, freedom of labor to join or not to join a union, and freedom of both employer and employee to bargain, either with or without participation of a labor union, comprise a second cardinal principle of such a program.

A third principle demands that protective tariffs on imports or exports be brought to a minimum, so that the restoration of free world trade may be promoted.

In a word, there is but one way to induce continuing growth of national income and employment; namely, continuous increase of the output of goods and services. This demands, above all, persistent reduction of the prices of those goods and services whose production is undergoing cost-saving through the application of new techniques and methods of production and through an increasing division of labor.

In a free economy cost-saving does not flow entirely, or even principally, into profits and wages; its benefits accrue to all economic groups through lower prices and increase of products. It was this passing on of the results of cost-savings to the public which, until the late 1920's, accounted for the rapid growth of the American nation. Continued progress along this line is dependent upon preservation of the free economy.

CHAPTER XXXII

Unemployment

Meaning of unemployment. In this chapter, we shall commence the study of certain problems confronting the wage earner in modern industry, selected for our discussion from a much larger number of labor problems because of their importance to society as a whole as well as to the individuals affected by them. The present chapter deals with the problem of unemployment.

The unemployed consist of workers unable to find employment though seeking it and competent to perform work. This definition does not include all who may be idle at a given time. Those who won't work—wage earners on strike and the work-shy—are not included.

It is a more difficult problem to decide whether the definition of unemployment should include the aged and disabled—people who have been wage earners and are now seeking jobs, but who, through accident or disease or because of the preference given to younger workers, have lost their ability to find employment. For the most part, these unfortunates are victims of our modern system of production. Their condition is frequently the result of the health hazards of industry or the tension and strain of present-day productive methods, which place a premium on youth. The statistics which we shall use in estimating the amount of unemployment include this group of people, but when dealing with the causes and remedies of unemployment it is well to keep them separate from the rest of the involuntarily idle, since in these respects the aged and disabled present a problem peculiar to themselves.

Extent of unemployment. Although complete and accurate information of the extent of unemployment is not available, its amount is known to be large even in times of normal business activity and to rise to astonishing proportions during depression periods. The accompanying table shows the course of unemployment since 1920, both in numbers and in percentages of the wage-

earning population. These estimates cover all types of employment except agricultural workers.

ESTIMATED UNEMPLOYMENT IN THE UNITED STATES[1]

Year	Average monthly number unemployed	Average percentage of all wage earners
1920	1,400,000	5.1
1921	4,300,000	15.3
1922	3,400,000	12.1
1923	1,500,000	5.2
1924	2,300,000	7.7
1925	1,800,000	5.7
1926	1,700,000	5.2
1927	2,100,000	6.3
1929	1,864,000	3.9
1930	4,735,000	9.7
1931	8,568,000	17.4
1932	12,870,000	25.9
1933	13,271,000	26.4
1934	11,424,000	22.5
1935	10,652,000	20.8
1936	9,395,000	18.2
1937	8,282,000	15.8
1938	10,836,000	20.5
1939	9,979,000	18.7
1940	9,104,000	16.9
1941	5,013,000	9.3
1942	2,382,000	4.4
1943	1,070,000	2.0
1944	842,000	1.6

Even in prosperous times there is a substantial amount of unemployment. From 1923 to 1929, nonwar years of active business conditions, there were never less than one and a half million workers involuntarily idle nor less than about 4 per cent of the total number of wage earners. Averaging the number of unemployed in good times and bad, the estimates indicate that some 8 per cent of all the workers are involuntarily idle all the time. So persistent is this condition that it has come to be called *normal* or *chronic* unemployment.

Unemployment becomes acute during times of depression. The

[1] Figures to 1927 from National Bureau of Economic Research, *Recent Economic Changes in the United States*. New York: McGraw-Hill Book Company, 1929, Vol. II, p. 478. Figures from 1929 through 1940 are estimates of the American Federation of Labor published in the *American Federationist*, January 1936 and May 1941. The year 1928 is not included by either of these sources. Figures from 1941 based on figures in War Manpower Commission, *Manpower Review*. Washington: Government Printing Office, June 1943 and June 1945.

disastrous unemployment which accompanies a prolonged depression is shown in the figures from 1931 to 1940 when the number of unemployed at its maximum in 1933 reached over 13,000,000 workers and represented over one out of every four wage earners. But even these figures, large as they are, do not tell the whole story, for the well-established practice of spreading work by placing the whole labor force on part time reduced the percentage of complete unemployment and threw a burden of partial idleness upon the workers who still held their jobs. At the end of 1937, when nearly 10,000,000 workers were wholly unemployed, an additional 5,000,000 were working on the average only 18 hours per week, or less than half-time.

When unemployment reaches such extremes, it becomes the major practical problem of the day. Through their effect on the employment market, the forces causing a recession of business activity renew themselves and tend to carry the depression to even lower levels. The initial maladjustment in the business world throws large numbers of men and women out of work. The cessation of their buying power in the market prolongs and increases the maladjustment, causing further recession of business activity and in turn additional unemployment. It is likely too that some of the things done by the government and other social agencies to relieve distress and make work for the idle act as independent causes retarding normal economic recovery or even aggravating the depression.

Causes of unemployment: Demand and supply of labor. The causes of unemployment in the broadest sense, however they differ in detail, have one essential characteristic in common. They are all aspects of maladjustment between the demand for labor and the supply of labor in the employment market.

The supply of labor, consisting of the services of human beings, is not as a total supply subject to sudden and substantial changes. The relative immobility of labor which impedes the transfer of workers from place to place and from one kind of occupation to another in response to changes in demand, though not a positive cause of unemployment, does have an aggravating effect upon the amount and duration of idleness resulting from other causes.

Since the supply of labor is not subject to sudden or capricious change, the forces which upset the equilibrium of demand and supply must operate principally on the demand side of the market.

There is an unexpected reduction in the need for wage labor or in the profitableness of its employment in some large occupational group, or in several such groups simultaneously, or in all industries together. Discovery of the major causes of these changes enables us to identify three principal types of unemployment, each with its own characteristics of cause and remedy: *technological, seasonal,* and *cyclical unemployment.*

Technological unemployment. The invention of new and the bettering of old machine equipment, reorganization of the layout of production within an enterprise, improved methods of management, and the like are salient features of modern industry. Any such change is considered an improvement only if it reduces costs of production. The saving of cost usually takes the form of a reduction of wages per unit of output, accomplished either by diminishing the number of workers employed or substituting a cheaper for a higher grade of labor. The immediate effect of technical progress therefore is to reduce the demand for the labor of some occupational group.

It is hardly necessary to present evidence regarding so commonplace a feature of our industrial order. Technological change is incessant in all lines of production. Generally the specific event has a relatively small effect on the labor market, but even minor changes, if they are real improvements, tend to unsettle the equilibrium of demand and supply at some point.

There are times, however, when technological progress seems suddenly to affect large areas of the industrial system and to exert far-reaching influences upon the wage-earning class as a whole. Workers in a given industry are not affected by changes in that industry alone, but may find the sale of their own product and hence the demand for their labor decreased by a reduction of costs in an enterprise producing a competing commodity. Thus the influence of technical advance is broadly diffused through the laboring population.

Although distorted in some respects by the depression of the 1930's, available figures of the changes in output, man-hours worked, and output per man-hour from 1923 to 1940 in manufacturing, transportation, public utility, and construction industries emphasize the cumulative effects of technological progress.[1] As com-

[1] Hagan and Kirkpatrick, "National Output at Full Employment," *American Economic Review.* Menasha, Sept., 1944.

pared with the base period, 1923–25, output by 1940 had risen 35 per cent while man-hours worked declined 19 per cent so that output per man-hour increased by 67 per cent. The decline in man-hours worked represents not only the shortening of the work week but also the temporary displacement of workers who suffered unemployment for various periods of time.

In contrast to the immediate effect, ultimately technological progress does not reduce and may actually increase the demand for labor. Increasing efficiency of industry offers an immediate opportunity to shorten the work week and thus counteracts the tendency to reduce the numbers employed. Of much greater significance, however, technological improvement, since its effect is to lower the costs and hence the prices of goods, constantly increases the volume of goods sold and promotes expansion of industry. Expanding industry creates a new demand for labor, which tends to reabsorb the temporarily displaced workers.

The amount of expansion and reabsorption of labor resulting from a given price decline will depend on the elasticity of demand for the good. If the demand is elastic it is possible that even more labor will ultimately be employed in the same industry after the change than before, although the increased demand may be for a different kind or grade of labor. If elasticity is comparatively slight there will probably be a permanent decline in the number of workers in that particular industry. But then consumers will be spending a smaller total amount of money on this particular good, and the purchasing power released by the change will be expended on other goods and services. Under either alternative therefore the final effect of the change on employment is not an absolute reduction of the number of workers employed but at most a shifting of demand from one labor group or one locality to another.

If other alternatives are examined it will be seen that these ultimate consequences remain the same. The new invention or scientific discovery may create an entirely new industry which undermines an older one, as the automobile has reduced the manufacture of horse-drawn carriages, or the radio has brought hard times to the vaudeville stage. But these developments, when they have worked out their effects, merely reduce the demand for one kind of labor and increase that of another. Or we may assume, as is sometimes argued, that the gain from a cost-reducing improvement

is not passed on to the consumer but retained by the owner of the industry. Even then the owner will normally spend his monopoly profits either through the investment market or for consumable goods, thus increasing demand for labor at some point in the productive system.

In the long run therefore improvement of production methods cannot permanently lessen the demand for labor as a whole; it merely redistributes the demand among the vocational groups which make up the wage-earning population. This however does not disprove the existence of an immediate problem. Each case of technological unemployment may be temporary in the long-run view of things, but it is nonetheless serious at the moment to the workers affected by it. It may even be permanently disastrous to certain groups whose crafts are rendered obsolete by new machines and whose means of livelihood on the former level of comfort have vanished.

Even for society as a whole the net result in eras of broad scale and continuous change is not temporary but prolonged unemployment. The individual cases add up to a large total of people out of work at any given time, and the unceasing march of technical progress replenishes the army of the unemployed year after year as different individuals drift in and out of its ranks.

Seasonal unemployment. Unequal distribution throughout the year either of the sale of the product of an enterprise or of the supply of raw materials it utilizes results in seasonal changes in the demand for labor. The causes of these fluctuations are so many and so various that no detailed account can be given of them here. In some cases it is habit and custom—fashion, the observance of holidays, etc.—which alter the demand of different seasons for the products of labor; in others climatic changes affect either the demand for some product, as clothing and fuel, or the supply, as in agriculture, or the facility with which production can be carried on, as in the building trades. Very few industries are entirely immune from seasonal influences, either directly or through the medium of other industries initially affected.

That these rhythmic fluctuations in the tempo of business activity create unemployment goes without saying. In the coal industry 40 per cent of the workers employed in the best month are idle in the worst; employment in the construction industry has

fluctuated as much as 79 per cent in the course of a year; in retail trade by 30 per cent; in clothing manufacture by 35 per cent; in automobile manufacture by 24 per cent.

It is characteristic of the irrevocably seasonal industries to attach to themselves a supply of labor equal to the demand at the highest seasonal point. Relatively high wages, bonus payments, overtime pay offered by employers to stimulate output at these peak points mislead workers into the belief that these vocations are unusually profitable. So there accumulates a *reserve of labor,* or a supply of workers in excess of the number that would be required to produce the same annual output on a stabilized plane of operations. In coal mining, for example, it was found at one time that 200,000 workers would be continuously idle if production were spread evenly throughout the year. This wasted labor is the ultimate social loss from the seasonality of industry. From the point of view of the worker the loss is self-evident.

Cyclical unemployment. The effect of the business cycle on employment varies with its different phases: an increasing demand for labor during the recovery phase, a relatively full and steady demand during the prosperity phase, declining demand during recession, abnormally small demand, more or less prolonged, during depression. How extreme may be the resulting unemployment in the trough of a depression is shown by the figures already given for 1921 and 1933. When conditions were at their worst, following the crisis of 1929, a quarter of American wage earners were out of work, many millions were destitute, and the combined resources of city, state, and national governments were strained to keep the people from want.

Depression ultimately finds its way into recovery, and cyclical unemployment gradually disappears. But recovery from a bad depression always brings with it far-reaching readjustments within the industrial structure, causing some forms of economic activity to shrink below previous levels, others to expand. The restoration of the demand for labor therefore may never reach some occupational groups, and many people may suffer unemployment prolonged far into the recovery period or be obliged to shift to new occupations or new localities to their enduring loss.

Attitude of the laborers. The fear of unemployment exerts a powerful influence upon the behavior of the wage-earning classes of society. Its causes are for the most part beyond their control. Yet

it is they who feel its effects most grievously, and it is but natural that they should seek to protect themselves by using any weapons which lie in their power.

When labor is organized in defense of its interests, the union is prone to adopt policies designed to prolong the job or to spread it over a larger number of workers. One such policy of union labor is the demand for what is known as the "full crew." Agreements between trade unions and employers in the building trades often specify that each group of craftsmen be accompanied by a minimum number of apprentices and helpers. Labor organizations on the railroads have demanded that each train be staffed by a certain number of workers of various classes, without much regard to the real utility of the full crew in the circumstances of the case. The seamen's unions have advanced similar demands with respect to ocean-going vessels.

Fear of unemployment is also in some measure responsible for the emphasis the unions place on seniority rules governing layoff and rehiring of workers. These rules and the disposition of employers to follow similar practices in the absence of union rules give security of employment to workers in accordance with their length of service. Workers therefore feel that they have a better chance of retaining a job if they have held it for some time, and this tends to hold them to their present employer even in the face of better wages elsewhere. It is not unlikely that this practice, stemming directly from the frequent occurrence of unemployment, is a major cause of the immobility of labor, which, as we have seen, tends to retard the correction of unemployment.

But the most far-reaching effect of the fear of unemployment upon the conduct of the workers is seen in the disposition of labor —the unorganized as well as the members of trade unions—to oppose measures designed to increase production. It is impossible to calculate how much society loses through the time-wasting devices of workers in fear of unemployment. Make-work tactics are not founded on any closely reasoned economic policy, but they do reflect a general belief that there is at any given time only a certain amount of work to be done. The problem of each laborer is to get his share of this quantum of employment. It is assumed that, if one man through extra exertion does the work of two ordinary workers, he drives someone else into unemployment; if all workers discharge functions with unusual efficiency, they will all be forced into idle-

ness a part of the time; employment is a thing to be husbanded, economized, made to last.

The importance of the job to the worker is so great that he develops a proprietary attitude toward it whenever conditions are favorable. Strongly organized groups which have entered into collective agreements with their employers fall into the habit of viewing the job as the joint property of employer and worker. They believe that neither party to the labor contract has the right to dispose of these jobs without consulting the other party.

An understanding of this view of the situation throws light upon the behavior of men when on strike. The strikers do not assume that their action has freed the employer from any obligation toward his former labor force and has authorized him to offer employment to anyone in search of work. They say they have quit work with the intent of retaining the job, that the strike is merely a contest over the terms of employment between the two parties who jointly control employment in the enterprise. Until this contest is settled, the strikers feel that the entrepreneur does not have a free hand in filling the vacant places; any worker who accepts employment during the strike is a job-stealer, a "scab", and against him aggressive measures may be employed with propriety.

Remedies for unemployment: Adjustment of labor supply: Mobility of labor. Among the remedies for unemployment we shall consider only those which, by attacking the causes of unemployment, show promise of working permanent improvement. Each type of unemployment calls for remedies pertinent to its own causes. But there are certain policies or programs of action which may be applied to the problem as a whole.

All types of unemployment result in the last analysis from the failure of the labor supply to adjust itself promptly to changes in demand. While steps may be taken to control the abruptness of these changes, it is obviously profitable also to adopt measures to reduce the lag in the response of labor to them.

This requires that wage earners be ready and able to desert localities or occupations whenever shifts in demand make them relatively overcrowded. A few moments' thought will disclose the obstacles which prevent this response. Wage earners are human beings, attached to their places of residence by those ties of sentiment, human relationships, and material interest which bind all social groups to their home communities. In the skilled trades the

individual's knowledge of the craft is a possession not applicable to other vocations and therefore lost when he changes his calling. Moreover willingness to make these sacrifices is not sufficient to enable the worker to move from one vocation to another, for it is often impossible for him to acquire the new skills demanded of him in the new line of work. Any policy which increases the worker's mobility in the geographic sense or his versatility in the craft sense will reduce the amount of unemployment.

Public employment offices. Frequently workers simply do not know that jobs are more plentiful in another locality or another vocation, and they may have no reliable sources of information on these matters. The employment market can be better organized, so that obstacles to the redistribution of labor due to mere ignorance of the facts will be removed. Organization of the markets for securities and the staple commodities has been perfected through exchanges which bring demand and supply continuously together.

The same principles can be applied advantageously to the labor market by means of the employment office and the labor exchange. A complete system of employment offices would cover the entire industrial area with an intercommunicating series of local agencies, each registering and classifying the idle workmen and the available jobs in its locality, and all coöperating to distribute unemployed labor over the entire area with as little delay as possible.

Most modern industrial nations, except the United States of America, have long had efficient nationwide systems of this sort under public management. In the United States until recently this service had been left to the individual states or to private profit-making agencies. A few of the states worked out fairly efficient systems, but for the most part the state employment office has been a perfunctory affair, staffed by political job holders. The private agencies, since they profit from a continuance of the evil they purport to remedy, are frequently even worse than this, adopting practices which partake of outright fraud.

The passage of the Wagner-Peyser act in June, 1933, marked the beginning of a new national policy with regard to this problem. This law created the present United States Employment Service with a central office in Washington, staffed with trained people under a director-general, who has the assistance of an advisory council representing employers and workers. To stimulate state action the law empowered the federal government to make an an-

nual appropriation of $4,000,000, of which $3,000,000 is allotted to the states according to population, on condition that each state which accepts the grant shall match the federal subsidy dollar for dollar and that the state's system of employment offices shall measure up to standards defined by the federal service. The remaining $1,000,000 is used to conduct an interstate clearing house of information, to inspect the state services, to carry on research and publication, and to render certain direct services to labor in the District of Columbia and among the war veterans.

At first the states were slow to respond to the invitation of the federal government to revamp their employment offices and affiliate with the United States Employment Service. By June, 1935, only twenty-five states had done so. But the passage of the Social Security Act in 1935 hastened the process. This law embraces a broad group of policies, but its provision of federal aid to state unemployment compensation systems, by necessitating the creation of state-wide employment offices to record the unemployed and distribute the compensation, has spread the public employment office system throughout the entire country. By 1938 all the states and territories had set up these systems, and all were affiliated with the United States Employment Service.

During its early years the service was to a large extent deflected from its normal long-range functions to discharge placement services required of it by various federal emergency agencies such as the CWA and the PWA. Its record nevertheless was impressive. From 1933 to 1938 some 26 million unemployed people were registered in the employment offices of the country as applicants for jobs; about 22 million placements were made, including those in the various work relief services.

From 1940 on, however, the number of placements made in private industry increased rapidly. During the war, the Employment Service, operating under the War Manpower Commission, had primary responsibility for directing available labor into war industries. During 1944 and 1945, it was required that all hiring in designated areas of acute labor shortage be done through the Employment Service or through channels approved by the service. As a result of these increased responsibilities, nonagricultural placements made by the Employment Service increased from about 3.7 million in 1940 to 11.5 million in 1944.

Flexibility of wage rates. As in the case of other goods, the behavior of the price of labor has much to do with the adjustment of supply to changes in demand. A falling wage rate in a locality or a vocation is at once the most accurate sign of oversupply and the most effective stimulus to its correction. A rigid wage rate set above the marginal productivity of a labor group will create and tend to prolong unemployment at that point in the market.

Wage rates are at best sluggish prices; their rise and fall is retarded by sentiment, by custom, by contract. Any arbitrary or mandatory policy which makes them still more rigid and unyielding must be viewed with suspicion. Such policies—legal minimum wages, code agreements pegging wage rates, and the like—are frequently advocated by well-intentioned people, usually in times of distress and always in respect of some labor group whose low or diminishing rate of pay reflects the oversupply of their services.

It may be desirable for moral or humanitarian reasons to adopt such policies on occasion, but this should always be done in full knowledge of their danger as a possible aggravation of the problem of unemployment. We may take it as a sound principle that anything that increases the plasticity of the wage rate will tend to prevent or diminish unemployment.

Stabilizing the demand for labor: Voluntary action by industry. Specific remedies for the different types of unemployment take the form of attempts to stabilize the demand for labor. These attempts may be made by employers acting voluntarily, or by the government acting independently outside the system of private industry. Voluntary programs of this sort require organization on both sides of the labor market, collaboration among competing employers, and collective action by the workers.

The interest of society as a whole in technological progress is so obvious and so paramount that no proposal to stabilize industry by putting an end to invention and improvement is worthy of consideration. But for this very reason—because the gain goes to society as a whole—it is now being recognized as neither just nor expedient that the cost should be borne solely by one social class.

Policies can be adopted by agreement between employers and workers which, while permitting technical advances, will slow down their rate and provide time to forestall unemployment. The linotype machine furnishes the classic example. A rapid wholesale

adoption of this invention in the printing trade would have destroyed an entire craft and caused much distress. By agreement it was introduced gradually, the former typesetters were trained to operate it, and the result was that no unemployment occurred even temporarily, while ultimately the demand for labor was greatly expanded through the decline of printing prices. The molding machine was introduced into the iron industry under similar controls. Very recently the textile workers and the national association of their employers have attempted by joint agreement to provide for more gradual adoption of the automatic loom.

The sharply seasonal industries are experimenting with various policies of stabilization in harmony with their differing technical conditions. The following is a summary of the more important of these devices:

(1) An industry whose major product is subject to fluctuating consumers' demand may take on *compensating side lines,* whose market varies inversely with that of the major product. An example is the manufacture of felt hats or cloth caps by a concern specializing in straw hats, or the production of some staple product by a concern devoted primarily to the manufacture of Christmas specialities.

(2) Somewhat similar is the development of *compensating markets.* Cement manufacturers have regularized their output by exploiting the South American markets where the seasons are the reverse of those at home. Much can be done on a smaller scale along this line within the continental area of the United States.

(3) Goods may be *made to stock* during the slack months. This device involves expense of storage and can be used with safety only in connection with types of product which do not deteriorate in storage and are not subject to capricious changes of fashion.

(4) Sales methods may be adjusted to encourage *advance orders* with a long delivery time. By offering particularly favorable terms, a large concern specializing in the Christmas trade has succeeded in inducing many of its customers to place their orders more than a year in advance, thus allowing a stabilizing of production. Similar results have been obtained by many of the larger clothing manufacturers.

(5) Companies which perform several different processes may set up a system of *vocational training,* increasing the versatility of their workers so that they may be transferred from slack to busy departments without loss of time.

Public works. Remedies for cyclical unemployment have been suggested in connection with our study of the business cycle and need not be treated in detail here. It is in relation to this phase of the unemployment problem that governmental activity is usually proposed. The program of public works which has figured so largely in the activities of the American national government in recent years is one example of this type of policy.

The public works program contemplates governmental activities on a large scale during periods of general business depression, involving works which would not otherwise be undertaken at all, such as slum clearance, reforestation, road building, irrigation projects. As advanced by serious students of the problem, the program is carefully safeguarded against the most obvious dangers involved in extravagant governmental spending. It is to consist entirely of projects worth while for their own sake; it is to be financed as much as possible from funds reserved for the purpose during prosperous years, thus avoiding the necessity of heavy taxation or large public borrowing during a time of unsettled business; it is to be administered in a businesslike way, employing laborers on the basis of their usefulness to the enterprise, not merely because they are unemployed, and holding them to the same standards of wages, working conditions, and performance demanded by efficient private enterprise.

In addition to these safeguards, a sensible public works program must be based on reliable statistics of unemployment to permit of advance planning in prosperous years and to regulate the scope and the regional location of the government's activity in harmony with the trend of unemployment. Assuming that all these features of the program can be assured, there is something to be said for such a program as a corrective of cyclical unemployment.

As things go in the political world there is grave reason to doubt that any such businesslike handling of so large and complex an undertaking could be obtained through the agencies of government. But even if wise administration were assured, it is still to be feared that a public works program would exert injurious indirect effects upon the situation. If funds reserved for public works in prosperous years are invested in income-bearing securities, grave difficulty may be encountered in liquidating these reserves when the money is needed, and the attempt to do so in a falling market may add to the difficulties of the business world.

The ultimate solution of business depression must be such readjustment within the price structure and correlatively within the system of production as will bring economic forces to a new point of equilibrium. With reference specifically to unemployment we have seen that anything that increases the rigidity of wage rates will impede solution of the problem. A gigantic public works program can hardly fail to create an artificial demand for certain types of goods and labor which will tend to reduce the flexibility of the price structure. To prevent this sort of unforeseen and indirect effect of governmental intervention in economic affairs is virtually impossible, even by the most careful administration.

The merits of an ideal public works program have never been put to the test of experience, for nowhere in the world has such a program been attempted. It is true that a vast program of public spending was undertaken by the United States Government as a remedy for unemployment and as a device to stimulate business recovery, beginning with the National Industrial Recovery Act of 1933, which appropriated 3⅓ billion dollars for this purpose. This and subsequent appropriations extending through 1938 were spent, in part as subsidies to state and local governments to assist in financing approved projects designed and administered by these governments, and in part on a number of large-scale enterprises undertaken by the federal government itself.

The program was under the immediate direction of the President and the Secretary of the Interior, who approved all projects proposed by the state and municipal authorities and saw to it, as a condition of federal subvention, that these projects received equivalent financial support from the local governmental units.

This experiment undoubtedly did much for the relief of distress and possibly exerted some influence in the business recovery of 1934–37. But however that may be, this experience did not constitute a real test of the public works method of solving unemployment.

The program was adopted as a substitute for an existing policy of unemployment relief, which was draining the resources of governments, local, state, and national. It had no such background of long-range planning as is proposed as a fundamental feature of the ideal public works program. For these reasons the labor employed was selected or managed, not by tests of efficiency, but as objects

of charitable relief, and the projects undertaken were devised on the spur of the moment in the midst of a great national emergency.

Moreover the program was financed, not out of reserves accumulated over previous years, but by current taxation and borrowing at a time when the investment market was already disturbed and the taxpayers already impoverished by the prolonged business depression. In all these important respects this activity of the government failed to qualify as a true test of long-range planning of public works.

Unemployment insurance: General nature and origin. The movement to protect workers as far as possible against the hazards of unemployment is supported by sound business considerations as well as by sentiments of justice and fair play. It seems manifestly unfair that a loss attributable to general social forces should be thrown upon the one class in society least able to bear it. Moreover it is a well-recognized fact that the fear of unemployment acts as a deterrent to the efficiency of workers, inducing them to adopt job-saving or time-wasting practices. Society would gain much by the removal of this fear as an offset to the cost of administering the policy of insurance. Finally, although insurance, strictly speaking, is only a means of distributing the loss, the policy may be so devised as to reduce the amount of unemployment, thus preserving for society a portion of the productive power which would otherwise be totally destroyed through the idleness of workers.

Certainly little progress can be made in the solution of unemployment as long as the problem itself is concealed from view. An insurance system could disclose the extent and location of unemployment, and its details could be so adjusted as to make it worth while for employers to keep their employees continuously at work. These considerations of justice and economy have caused most modern nations to adopt some form of unemployment insurance.

Unemployment insurance began with the efforts of the trade unions to provide benefits for their members when out of work. The procedure followed is simple. Membership dues are set at a figure large enough to accumulate a fund out of which to pay benefits to the unemployed. As a safeguard against fraudulent claims the benefits are disbursed by local officials who know the applicant personally, and maximum limits are set to the duration of relief in individual cases.

These voluntary systems are to be found in practical opera-

tion in all countries where trade unionism has attained strength. The contributions made by the members are not based upon actuarial data; the reserves set aside against future benefits are inadequate to meet the drain of widespread unemployment, and accordingly the systems collapse when the need is greatest; at best only a minority of the working class (*i.e.*, the part which is organized in unions) is benefited. Furthermore, since employers bear none of the cost of relief, trade union insurance adds nothing to their incentive to keep their workmen employed.

The Ghent system. Public insurance began as a half-way measure, taking the form of subsidies from the government treasury to support systems of trade union insurance. This combination of mutual and public insurance has come to be known as the *Ghent system,* taking its name from the city which first adopted the policy. After its beginning in 1901 the system spread widely among the municipalities and larger governmental units of Europe, until in 1935 it had become a well-established policy in Belgium, Czechoslovakia, Denmark, Finland, France, Holland, Norway, Spain, and Switzerland.

In all these countries the policy had similar features. Public contributions were made on condition that a substantial portion of the cost—usually one-third or one-half—be borne by the trade union; administration was placed in the hands of union officials; certain standards determining the amount and duration of benefits were set by the contributing governmental body.

The Ghent system is superior to simple trade union insurance because of its greater financial adequacy, but otherwise it has all the weaknesses of that type of insurance mentioned above. Furthermore it raises serious problems when applied in countries where employers oppose independent trade unionism, for by basing its policy upon the unions, the state will assist them to enlarge their membership and so will appear to be taking sides in the controversy. The failure of voluntary methods to provide adequate relief for the ever growing burden of unemployment has led many governments to pass laws making insurance compulsory.

Compulsory insurance in Great Britain. Great Britain was the first important country to adopt compulsory unemployment insurance and the one country which has given the policy thorough trial on a comprehensive scale and over a long period of time. The original law passed in 1911 applied to only a few selected trades, in-

cluding a small fraction of the wage earners. The coverage has been broadened at various times since that date so that today most of the working population is embraced in the system.

Employers, workers, and the Treasury contribute equally to the Unemployment Fund, the contributions of the workers being held back by the employers and passed on to the government. The payment of benefits is carefully safeguarded. To be eligible for benefits the unemployed worker must have made at least thirty payments during the preceding two years. He must register at an employment office and is obliged to accept any suitable job offered him. There is a waiting period before benefit payments begin, and the duration of the benefit period is strictly limited.

After ten years of successful operation this insurance system ran into increasing difficulties during the 1920's, which resulted in its virtual collapse. This was a period of severe business depression, with widespread unemployment and steadily increasing distress among the laboring population. Motives of compassion caused the authorities to relax the safeguards and restrictions with which the system was surrounded. The waiting period was virtually extinguished; eligibility was no longer based on the worker's contributions during the preceding two years; payment of benefits was continued after the statutory period had passed and the worker had exhausted his legal claim. In other words, the system ceased to be one of insurance and became a policy of charity to people in distress administered on a nationwide scale.

Before the end of the decade the insurance fund had exhausted its reserves and was heavily in debt to the government, which found itself carrying an impossibly heavy financial load. In 1931 the government, against violent political opposition, introduced drastic reforms, which restored the previous safeguards and transferred to the charitable agencies the care of those who were ineligible for, or had exhausted their rights to, insurance benefits.

The British experience discloses both the merits and the weaknesses of compulsory unemployment insurance. It has proved that in times of fairly stable industrial conditions unemployment is an insurable risk susceptible of treatment in a businesslike way. Under the proper safeguards it does not promote malingering, as has been feared, and it does act to reduce unemployment by stabilizing the demand for labor. But it also shows how vulnerable are these safeguards, particularly in times of widespread distress, and how readily

the system may give way to the demand that insurance be turned into a policy of charitable relief.

It seems clear that unemployment insurance will not stand the strain of severe business depression. Even if the political dangers of such times be avoided, cyclical unemployment cannot be handled by insurance methods. The cost will outrun any reserves that can be accumulated in advance. Even if enormous reserves are accumulated during long periods of good times, the attempt to liquidate them in an unsettled money market must tend toward further depression of business conditions.

Unemployment insurance in the United States. Under the American constitution the policy of unemployment insurance lies within the province of the several states. The federal government may take steps to encourage the adoption of state-wide systems by offering subsidies out of the national treasury, and if these subsides are accepted it may assume supervison of the standards and methods employed in the state plans. The Social Security Act of 1935 makes use of these federal powers.

This law embraces a broad program of social legislation, including provisions for old age pensions, care of the blind and crippled, maternal care and child welfare, as well as compensation for the unemployed. To handle the new insurance and benefit programs, the Social Security Board was created, consisting of three members with headquarters and a large staff in Washington and in twelve regional offices in various parts of the country.

The law authorizes $4,000,000 during the first fiscal year and $49,000,000 in each succeeding year to be allocated to the states as a contribution toward the administrative cost of such insurance systems as they may set up. As an additional inducement to the states, the act imposes a tax on the pay rolls of every business employing eight or more workers, at rates rising to 3 per cent in the third and each succeeding year, and provides that the employer may offset against this tax 90 per cent of any payments he makes to a compulsory insurance fund under his state law.

The law allows the states complete discretion in setting up systems of insurance suited to their own requirements or rejecting the policy outright, but requires that to qualify for grants from the federal fund the state system must meet certain tests with regard to its administrative and benefit features, that funds accumulated by the states must be deposited in the federal Unemployment Trust

Fund of the United States Treasury to be invested in federal securities, and that the state system must report regularly to the Social Security Board and hold its records continuously open to the board's inspection. The act was declared constitutional by the Supreme Court in 1937.

At the present time the forty-eight states, the District of Columbia, Alaska, and Hawaii have adopted approved unemployment insurance laws. The different systems exhibit such great variety in matters of detail that it is impossible to do more than summarize their chief features.

In all cases certain groups of workers are excluded by definition: agricultural labor, domestic servants, public employees, employees of religious, charitable, and educational institutions, and some others of minor importance. Moreover in all but ten states workers in small concerns—usually defined as consisting of less than eight employees—are not insured. It is estimated that some twenty million wage earners in all are covered by the insurance systems.

In all but four states, where the workers are required to make small contributions, the cost is borne by the employers in the form of payroll taxes. All the laws require a waiting period after the loss of employment before the benefits begin; all set a maximum period, frequently sixteen weeks, on the duration of benefits, requiring a minimum period of re-employment before additional benefits will be paid; and all adjust the benefits to the previous rate of earnings with a stated maximum, usually between $15 and $20 a week.

The state unemployment compensation laws became effective during 1938 and 1939, and most of them have since been amended several times. The general tendency has been to decrease the income of the funds by adopting experience rating systems so constructed as to reduce the average payroll tax on employers, and to increase the outgo of the funds by reducing the waiting period, lengthening the benefit period, and increasing the benefit rates.

These tendencies have thus far created no serious fiscal problem, owing to the rapid increase in employment after 1939. During the war years the yield of the payroll taxes increased greatly, while payments out of the funds fell to negligible proportions. The balance in all of the state unemployment compensation funds increased from 1.28 billion dollars on June 30, 1939, to 7.87 billion dollars on June 30, 1947. The funds proved amply sufficient to provide for

the moderate postwar unemployment which developed during the winter of 1945–46. Whether the present tax and benefit rates will prove actuarially sound over a longer period depends on the amount of unemployment which may occur in the future.

Unemployment compensation systems can provide a limited income to assist regularly employed workers through short periods of unemployment. Workers who are unemployed so long that they exhaust their benefit rights, or workers who work so irregularly that they cannot qualify for benefits, must look to some other means of support.

In the United States, resort must usually be had to cash relief administered by local governments and financed from local and state funds. The scale of relief payments varies greatly in different parts of the country. In general, payments are much lower than those provided under unemployment compensation, are adjusted to the size of the family unit, and are provided only after proof that the family has no other resources.

CHAPTER XXXIII

Other Labor Problems

Hours of labor: The shorter workday and work week. In industry as a whole, a strong trend toward the shorter workday became apparent over a century ago and has continued down to the present time. The shortening of the work day has been accompanied in recent years by reducing the number of workdays in a normal week. The effect of this movement in America upon normal hours of labor in succeeding decades from 1840 to 1945 is shown in the following table:

NORMAL HOURS IN MANUFACTURING AND MECHANICAL ESTABLISHMENTS, UNITED STATES[1]

Year	Average normal hours per day	Average normal hours per week
1840	11.4	
1850	11.5	
1860	11.0	
1870	10.5	
1880	10.3	58
1890	10.0	57
1900	9.8	55
1910	9.4	50
1920	8.5	49
1930	8.3	45
1940	—	38
1944	—	45
1945 (Oct.)	—	41

The depression following 1929 produced an abnormal situation as regards hours of labor, since many of the policies adopted by individual employers or governmental bodies to combat unemployment were of the share-the-work variety, involving part-time employment for many workers. In response to these influences, there

[1] Figures through 1930 from Daugherty, C. R., *Labor Problems in American Industry*, Revised edition. Boston: Houghton Mifflin Company, 1938, p. 192. Subsequent figures from *Monthly Labor Review*. Washington: Government Printing Office.

has been a strong movement toward a five-day week with not more than eight hours per day. In 1931 less than 800 concerns, with a total labor force of some 700,000, were known to be operating on the five-day week schedule; in 1935, 250,000 firms, employing 12,000,000, had adopted this schedule. The increase of weekly hours between 1940 and 1944 was due, of course, to the pressure of war. Following the end of the war in 1945, the five-day week was widely adopted by American business.

Hours and productivity. The major problem presented by this tendency toward shorter hours is the effect on wages. When workmen demand a shorter day they expect the rate of wages to be raised simultaneously so that their earnings may remain unchanged. Now it has been proved historically that earnings can be maintained or even raised while hours are reduced. During most of the period covered by the preceding table real wages rose steadily while hours of labor were declining. To what extent this can continue must depend in the final analysis upon the effect of the shorter workday on the productivity of the workers.

Within limits a reduction of hours has been found to increase the productivity of labor. It has been proved, by the experience of many different enterprises in Europe and America, and established as a rule of wide applicability by researches carried on by the British government and by various bureaus of investigation in America, that a shortening of the more extreme workday, a fall of hours from twelve to eleven or ten for example, results in an actual increase in the total output of labor.[1]

To be sure, this effect of shortening the long workday is not uniform for all grades of workers. It is most noticeable in those occupations which demand continuous mental alertness combined with manual dexterity, and is least so of the automatic types, such as the tending of machines whose pace is governed by the flow of steam or electrical power. But even the most unskilled labor responds to a reduction of an extremely long day by some increase of output.

There are many reasons for this. When the length of the workday is excessive, the period of rest is not sufficient to counteract the effects of fatigue. In extreme cases accumulated fatigue predisposes

[1] The evidence in support of this conclusion is summarized in the *International Labour Review*, International Labour Office, Washington, D. C., for April and November, 1924. See also Vernon, H. M., *Industrial Fatigue and Efficiency*, London: G. Rutledge and Sons, Ltd., 1921.

the workman to sickness, which results in lost time, and where this consequence is avoided the workman tends to be slack in attendance through voluntary idleness. It has been found that the accident rate varies directly with the length of the workday, reaching a maximum during the late hours, and also that it rises with the days of the week, thus disclosing another waste due to fatigue. Waste of materials in the form of defective output, waste of power, and necessity of more labor of superintendence through increased slackness of the workers also result from protracted hours of labor.

A saving of these wastes can be effected by shortening the hours of work to avoid the worst effects of fatigue. It is probable that the ten-hour day is more productive at the present time for all grades of labor than is the twelve-hour day and that for the most skilled groups an eight-hour day is at least as productive as ten hours of work.

But eventually the shortening of hours of work must strike a point where productivity of labor is not increased but diminished, unless this effect is forestalled by advances in the technique of production. Given continuous technological advance, the workers may choose between two alternative ways of taking their share in the general improvement. They may elect to receive their former rate of real income and take their gain in shorter hours of work or, working the same number of hours as before, to raise their standard of material comfort.

During the past decades their actual choice has taken a middle ground between these extremes. As their bargaining power has increased, the workers have pressed for a shorter workday and for a simultaneous increase of wages. That they have won both concessions is to be attributed, in part to the positive increase of productivity resulting from reduction of long hours which we have discussed, in part also to the fact that these demands were made at a time when industrial progress was working in the direction of increasing the hourly productivity of labor.

For some groups of workers the shortening of the workday has maintained their productivity at its maximum; for others it has resulted in an output below the maximum of which they are capable at the present stage of the industrial arts but higher than was possible at earlier stages.

Child labor: Economic characteristics. The labor of women and children is not a phenomenon peculiar to the modern system of in-

dustry. It is characteristic of all types of societies at all times. That it should be viewed today as a social problem is due chiefly to the fact that gainful labor by these members of the population has economic consequences in a society like our own which did not result from it in former times. We shall not stress the moral aspects of the problem, or the health hazards encountered by women and children in industry.

There are also various economic consequences which need only be mentioned, such as inability of the child prematurely employed to acquire the skill demanded of the higher wage group; the likelihood that he will be lodged inescapably in a blind alley occupation; the resulting maldistribution of the labor supply as between the overcrowded and the undersupplied occupations. Similar problems arise in connection with the employment of women.

Extent of child labor. American official statistics of child labor do not exist for the period prior to 1870. The United States Census of that year reported 739,164 children between the ages of ten and fifteen at work. This number increased to a peak of almost 2 million in 1910. The marked decline in child labor since 1910 is shown in the following table, which includes all children aged ten through seventeen.

CHILDREN AGED TEN TO SEVENTEEN IN THE LABOR FORCE
1910–1940[1]

Year	Number		Percentage of all children in age group	
	Male	Female	Male	Female
1910	2,262,000	995,000	31.0	13.8
1920	1,859,000	986,000	22.7	12.1
1930	1,473,000	755,000	15.4	8.0
1940	940,000	386,000	9.8	4.1

The downward trend was interrupted briefly during the Second World War, when large numbers of young people entered employment prematurely because of an abundance of jobs at attractive wage rates. It is estimated that the number of workers (including members of the armed forces) aged fourteen through seventeen

[1] Adapted from Long, Clarence D., *The Labor Force in Wartime America*, New York: National Bureau of Economic Research, 1944, Table 1. The figure for 1940 includes only children aged 14–17; no data were collected by the census on workers aged 10–13 on the assumption that their number was negligible.

tripled between 1940 and 1945, reaching a peak of almost 5 million. Since the end of the war the number has decreased greatly, and it seems reasonable to expect a return to the long-term trend.

Agriculture has always been the chief employment of child labor. In 1940, 48.3 per cent of all children aged ten through seventeen in the labor force were at work in agriculture, as compared with 11.7 per cent in manufacture, 1.4 per cent in transportation, 12.4 per cent in trade, 11.0 per cent in domestic service, and 2.2 per cent in clerical occupations.

The very recent trend of child labor is rather difficult to measure. Under the codes adopted in accordance with the provisions of the National Industrial Recovery Act of 1933, the employment of children under sixteen was universally prohibited in industry, while some hazardous lines of occupation set the minimum age at eighteen years. There was an abrupt decline in the number of child wage earners, estimated at two-thirds of the former total. A federal law of 1938, discussed below, has continued the restriction of child labor begun under these codes.

Effects of child labor. The effects of child labor do not require elaboration. The injurious moral effects which flow from an imperfect education, a denial of the period of play and recreation indispensable to mental and physical health, and contact with an environment which is often contaminating are apparent to all. There are as well certain economic effects which react unfavorably on the material welfare of society at large.

The wage-earning child is restricted in his choice of occupation to lines in which there is no hope of considerable advancement in earning power. Even when he is not physically crippled by overstrain or by abbreviation of the normal period of rest, he is denied opportunity to acquire the skill demanded of those in the higher wage groups. At maturity he is usually caught in a "blind alley" occupation, or recruited to the ranks of the already overcrowded group of unskilled manual labor.

This involves a loss, not only to the victim of child labor, but to all the rest of society's members. For as we have seen in connection with our study of differences in wages, the total income of society would be increased if it were possible to transfer men from the relatively overcrowded groups whose marginal productivity is low to the less crowded occupations where marginal productivity is higher. Any device which hinders this transfer not only tends to

perpetuate the inequalities of income as between the different classes in society, but also prevents a possible net increase in the total income fund.

Legal regulation of the labor of children. To the extent that it is desirable to prevent or regulate child labor, the necessary steps to this end must be taken under governmental compulsion. The problem is one which will not be solved by the free play of economic forces. It is undoubtedly true that many parents are eager to get income from the labor of their children at as early an age as possible. The children too are often eager to go to work as an alternative to going to school. Employers in a competitive industry adaptable to the labor of children can scarcely refuse, as individuals, to employ them, lest the competition of the less scrupulous competitors drive the others out of business. Here then is an appropriate field for public regulation.

All modern states have attempted by law to prohibit the worst abuses and to set standards in protection of children allowed to work. A comparison of the first child labor law of Great Britain with the present model code of that country will show how ideas of what is appropriate have changed in the course of time. This first law (1802) applied only to pauper children; it was considered a violation of individual rights to legislate for others. The law prohibited the labor in factories of paupers below the age of nine; children over nine were allowed to work twelve hours a day.

By the law of 1920 fourteen years is made the minimum age for industrial employment for all children; the length of the work day for young persons between sixteen and eighteen years is limited to eight hours; no work is allowed at night for any wage earner below the age of eighteen. Other European nations of advanced industrial civilization have approximately the same standards. These laws are sanctioned by heavy penalties and administered by an efficient system of government inspection, which makes evasion extremely difficult.

The regulation of child labor in the United States was until 1938 entirely in the hands of the states, and practically every state has a law on the subject. Twenty-eight states set a minimum of fourteen years; four, a minimum of fifteen years; fifteen, of sixteen years. Forty states set a limit of eight hours to the working day for children under sixteen. Only one state fails to prohibit night work for children, either entirely or in selected industrial occupa-

tions. The laws do not apply to work in agriculture or domestic service, and only a few of the states have attempted to regulate the street trades of children.

The practical effect of these laws varies, not only because of their different standards, but also because of the wide difference in the rigor with which they are administered in different states. In general, the southern states, where the factory system is of late development and where a necessitous population with low living standards creates a large supply of child workers, have the lower standards of regulation.

Lack of uniformity has been found to affect injuriously those units of a competitive industry which are located in the states having the more rigid standards of regulation, and this fact has tended to arouse opposition to child labor laws in these advanced states. Consequently various attempts have been made to bestow on the federal government power to impose uniform standards upon the country as a whole. A federal law passed in 1916 attempted to prohibit the labor of children under fourteen, to limit the workday of those from fourteen to sixteen, and to prohibit the labor of children at night, by excluding from interstate commerce the products of industries violating these standards. When this law was declared unconstitutional in 1918, Congress passed another statute embodying the same standards and attempting to regulate child labor by applying prohibitive taxes to the concerns violating the standards. In 1922 this law also was judged unconstitutional. Another attempt to give the federal government authority over child labor took the form of an amendment to the Constitution adopted by Congress and submitted to the states for ratification in 1924. This amendment would bestow on Congress power to prohibit or regulate the labor of children under eighteen years of age. In this case, however, ratification by the necessary majority of the states has not been obtained.

The latest attempt to prohibit child labor by federal statute was embodied in the Fair Labor Standards Act of 1938. With numerous exceptions, industries engaged in interstate commerce are prohibited by this act from employing children under sixteen years of age, while in hazardous industries the age limit is set at eighteen years. This act was declared constitutional by the Supreme Court in June, 1938, and has become an effective barrier against wholesale employment of children under sixteen in manufacturing industries.

Labor of women: Extent and character in the United States. The rapid increase in the number of wage-earning women in the United States during recent years is shown by the following table:

GAINFULLY EMPLOYED WOMEN AGED TEN YEARS AND OVER
1880–1940[1]

Year	Number	Percentage of all women of age group	Percentage of all employed persons
1880	2,647,000	14.7	15.2
1890	4,005,000	17.4	17.1
1900	5,319,000	18.8	18.3
1910	8,076,000	23.4	21.2
1920	8,550,000	21.1	20.5
1930	10,752,000	22.0	22.0
1940	12,845,259	25.4	24.3

For the period as a whole the aggregate number of women wage earners increased almost fivefold, outstripping by far the rate of increase of men workers. This increase continued at a rapid pace up to 1910, after which year the expansion of numbers, though continuous, was at a diminishing rate as shown by the percentages in the final column of the table. In 1940 almost a fourth of all wage earners in this country were women. During the war, because of the withdrawal of some twelve million men for the armed forces and the entrance of four million women into the labor market, the proportion of women workers rose to one third. While the proportion is now decreasing, it will probably remain permanently above prewar levels.

An examination of the distribution of women workers among the various occupations shows that there are now few types of employment which women do not share with men. If the occupations calling for special muscular powers are left out of account, all other broad divisions of occupations will be found to employ women in large numbers. An analysis of the 1940 census of occupations gives the information on this point, shown in the table on the opposite page.

Despite this widespread distribution, competition for employment between women and men is an important factor in only a relatively small number of occupations. If the major groupings of occupation

[1] Figures through 1930 from Bureau of the Census, *Census of Occupations*, Abstract Summary, 1932, pp. 3–5. 1940 figures from Bureau of the Census, *Statistical Abstract of the United States, 1947*. Washington: Government Printing Office, 1947, p. 169.

given in the table were broken down into a more refined classification, as in the case of clothing factories and textile mills, this concentration of competition would be clearly shown.

PER CENT DISTRIBUTION BY SEX AMONG THE PRINCIPAL OCCUPATIONS[1]

Occupation	Male	Female
All occupations	75.8	24.2
Agriculture	94.3	5.7
Manufacturing, total	77.7	22.3
Clothing factories	31.8	68.2
Textile mills	59.8	40.2
Trade	72.9	27.1
Professional service	44.2	55.8
Domestic service	11.9	88.1
Clerical	48.6	51.4
Sales	72.8	27.2

Causes of the increase in woman labor. To some extent, the greater employment of women is due to the progress of machine technique. But the movement is not wholly accounted for by reference to this factor. Mechanical equipment is of chief importance only in manufacturing, where, as we have seen, the labor of women has not risen to such formidable proportions as in other lines of occupation.

For a complete explanation of the increase we must look to tendencies of broad social significance in the customs and habits of the people. The postponement of marriage, the removal of customary restraints which formerly circumscribed the status of women in the home and the neighborhood, and the increase in amount and change in character of the education offered to women are all forces working toward an increase in the disposition of women to seek positions of economic independence. Founded as it is on such fundamental social forces, this tendency toward increased employment of women is probably irresistible.

There are unquestionable gains resulting from the influx of women into industrial occupations. Society has gained materially, through the substitution of productive for meaningless tasks for its unmarried women, and spiritually, through the release of energies formerly thwarted by the bonds of custom and tradition.

But these benefits have not been unmixed with other consequences. There are obvious moral and health hazards, and there are

[1] From Bureau of the Census, *Statistical Abstract of the United States*, 1940. Washington: Government Printing Office, 1941, p. 143; based on 1940 census of the labor force.

far-reaching influences upon the home—still the basic unit of our social structure.

Aside from the minimum wage laws of certain states, which are discussed below, laws bearing on the labor of women have been devoted primarily to setting a maximum limit on their hours and to prohibiting or limiting woman labor in certain unwholesome employments. At the present time, only four states fail to set some sort of legal limit to the workday for women. Fifteen of the states set the maximum at forty-eight hours a week; in eleven states the maximum is fifty-four hours; a few allow a work week as long as sixty hours. Twenty states forbid the employment of women at night, and many have special laws safeguarding the health of the woman worker. Little is done to regulate night work. The laws specifically exempt farm work and domestic work from regulation.

Wages of women. The feature which makes child and woman labor—particularly the latter—a distinctively modern problem is the effect of their competition on the welfare of men workers, tending as it does to undermine wage standards throughout the employment market. All the studies of this question concur in the finding that women are paid less than men for the same task and in the same industries.

One investigator, comparing the levels of wages in 1899 and again in 1923, discovered that in both years women's annual earnings were 54 per cent of men's in comparable employments. The National Industrial Conference Board found this same ratio in 1929 between the wages of women and men in skilled employments, while in unskilled occupations women received in that year 72 per cent as much as men. The best available information indicates that the differential between men's and women's wages in 1937 was about the same as in 1929. During the war the wages of women compared more favorably with those of men, but it is not unlikely that the earlier condition will be reestablished in postwar years.

Some part of this discrepancy is to be explained by the inferior productivity of the woman worker, but when all possible allowance has been made for this factor, there still remains a discrepancy between the wages of men and of women.

Systematic underpayment of women beyond what is due to inferior productivity cannot be accounted for except on the ground of their weaker bargaining power. As compared with men, women are notably weak in asserting their claims against the employer. We

have seen that they are concentrated for the most part in certain lines of occupation. Here they meet severe competition from a fluctuating group of women laborers who seek part-time or temporary employment as a supplement to their labor at home. Again they are even more immobile than the men; it is difficult for them to play off one employer against another when this involves a change of residence. In many cases also they are not self-supporting and make no pretense of demanding a "living wage"; the minimum wage which they will accept is lower than that of men, because they are partially subsidized by the male wage earners of their families. Those young women who look upon their employment as a temporary expedient to fill in the time before their marriage are not always intent upon exacting the highest possible price for their labor. Finally, they have been until lately unorganized and are still to some extent unorganizable; their bargains are made as individuals, and they labor under all the handicaps which the isolated worker must bear in the modern labor market.

Men workers look upon the competition of poorly paid women as destructive of their own standards of wages, as indeed it is. In their struggle to protect themselves against this situation, the men workers have devised the slogan so frequently found in trade union literature: "Equal pay for equal work." Two very different meanings may be attached to this demand that the wage rate for women be the same as that for men.

In the first place, it may mean that the payment for a given product should not be less for women, solely because they are women, than it is for men. Since women are admittedly less productive in most occupations than men, the equal pay slogan in this sense would result in generally lower day wages for women.

The second meaning resolves the slogan into a demand that the daily or weekly wage of women be the same as that of men, regardless of the relative productivity of the two groups of workers. If applied in this sense, the result would be generally higher wages in relation to productivity for women than for men. As advanced by most of the labor unions, the demand for equal pay has the second of these meanings. Its intention is not to raise the wages of women, but to exclude them from employment.

Minimum wage laws: General summary. Belief that wages too low to provide a decent standard of living are immoral or socially inexpedient has led governments to pass laws prescribing minimums

below which wage rates must not fall. Beginning with the legislation of New Zealand in 1894, this movement has spread through many countries until today sixteen of the principal industrial nations—among them Great Britain, France, Germany, Austria, and the British Dominions—have statutes of this character.

In the United States this form of social legislation has been in the hands of the several states except for the brief period 1933–1935, when the wage codes of the National Industrial Recovery Act were in effect. A Supreme Court decision in May, 1935, which held this act of the federal government unconstitutional, again gave the states sole jurisdiction over the matter. After ruling variously on the constitutionality of state minimum wage laws at different times, the Supreme Court declared such laws valid in March, 1937. The spread of state laws, which had previously been retarded by doubt as to their constitutionality, began again. In 1945, twenty-six states had adopted minimum wage laws, applicable (except in two states) to women and children only.

The United States Fair Labor Standards Act of 1938. The latest attempt of the United States federal government to establish minimum wages on a nation-wide scale took the form of a law bearing the above title, which was approved on June 25, 1938. The law applies only to industries engaged in interstate commerce and within this broad classification exempts many specific occupations from its provisions. For example, seamen, fishermen, agricultural workers, workers engaged in processing milk and other perishable foods, and streetcar and bus-line operators are exempt.

In the industries to which it is applicable, the law prescribes maximum weekly hours as well as minimum hourly wages. It also prohibits the employment of children under sixteen years of age in these industries and sets the minimum age at eighteen years if the industry is classified as hazardous. The provisions for maximum hours are as follows: 44 hours per week during the first year (*i.e.*, until November, 1939); 42 hours per week during the second year; 40 hours per week thereafter.

The basic minimum wages set by the Fair Labor Standards Act are 25 cents per hour during the first year of its operation, 30 cents per hour during the next six years, and 40 cents per hour thereafter. The immediate effect of these rates was to bring about wage increases for several million workers, particularly in the textile and other light manufacturing industries employing principally women

and young people. The minimum wage and overtime pay provisions of the act appear to have been reasonably well enforced. Between October, 1938, and June, 1943, some 110,000 employers were ordered to restore about 85 million dollars of illegally withheld wages to some 2½ million workers.

After 1941 the importance of the act diminished because of the rapid increase of wage rates already noted. Competition for labor arising in the war industries forced wages to rise, and the National War Labor Board, which was responsible for controlling wage increases, gave virtually automatic approval to increases up to 50 cents per hour. By the end of the war very few workers in interstate industries were receiving less than 50 cents an hour, and the 40-cent minimum set by the law acted only as a potential check on wage reductions. Hourly rates continued to rise in 1946, 1947, and 1948.

Appraisal of minimum wage laws. Minimum wage legislation presents an interesting problem in economic theory. As we have seen, the economic law of wages holds that the wage rate of any homogeneous group of workers will tend to settle under competitive conditions at the point of their marginal productivity. Advocates of the "living wage" and similar standards for the payment of workers challenge the validity of this economic principle, pointing to the successful operation of minimum wage laws in many countries as proof that the wage rate can in fact be set in accordance with standards of justice or social welfare. On the other hand, others have sometimes appealed to the law of wages to prove the futility of minimum wage legislation.

As a matter of fact neither of these extreme positions is valid. The principle of marginal productivity does not deny that wages can be set by statute; it merely implies that, if wages so determined are raised to an artificial level above the marginal productivity of the group, the result will be unemployment for some of the workers affected by the law. The government can require employers to pay wages at a certain rate if workers are hired at all; it cannot force the employer to hire any given number at this rate. If the legal wage is artificially high, the employer can and will equate the wage with marginal productivity by reducing the number employed.

The facts of experience argue either one or both of two things. First, it may be that the competitive wage in operation before the passage of the law was lower than the economic wage or, in other

words, that the bargaining weakness of the workers had been exploited to the end that they were paid less than they were worth. Then the actual wage rate can be raised by any device which strengthens bargaining power. Secondly, workers of low productivity may have shown an increase in efficiency, in response to better living conditions and the greater contentment which followed upon a rise in their standard of life. This factor too has been discussed in the preceding pages. The initial rise of wages which produces this improvement in productive efficiency may come through legal enactment, but if the ultimate result is a rise in the marginal productivity of the group, it is evident that the higher wage rate may become permanent, and that with no reduction in the numbers employed.

Whether the minimum wage will prove a real boon to labor or an ultimate loss depends then upon the caution with which the policy is applied and the extent of improvement which is attempted through it. Examination of the laws of the American states and of England, which have succeeded in raising wages without creating unemployment, shows that these limitations have generally been carefully observed. The usual method of determining the legal minimum has been to take as standard the wage rates of the more generous employers in a given occupational group. The statute in most cases makes this policy general throughout the industry, enforcing it against those employers who have been prone to use their bargaining power to exploit the workers.

Security for the aged: General. It has been found that nearly two-thirds of the population of the United States who are sixty-five years of age or older are dependent upon others for their support. A similar condition prevails in all industrial countries and has induced modern governments to adopt various measures to provide economic security for aged people.

These policies are of two kinds: *pension systems,* the provision of regular monthly grants to the impoverished aged from the current revenues of the public treasury, and *insurance,* a system based on actuarial principles and providing annuities for the aged regardless of their need, covering the cost by collecting contributions during the productive lifetime of the beneficiary.

Old age pensions. Immediately before the Second World War old-age pension systems were in operation in about a dozen countries, including France, Great Britain, and the United States. In the

latter country the policy was in the hands of the several states. Definite provision for old-age pensions by state law began in the twenties; by 1935 twenty-eight states had such laws in operation, paying pensions totaling upwards of $30,000,000 to nearly a quarter million people. The plans all contained about the same features: the eligible age was sixty-five years; American citizenship and five years' residence in the state were required; the amounts paid were usually limited to $30 a month; payments were made only on proof of need. This type of provision for the aged poor, developing slowly throughout the country, was given an immense impetus by the federal Social Security Act of 1935.

The Social Security Act provides for old age insurance as well as for assistance to the pension plans of the states. The latter feature is covered by a section of the act empowering Congress to appropriate $49,750,000 in 1935–36 and in following years as much as may be needed to make grants to states which set up old age pension laws approved by the Social Security Board.

The approval feature has been used to strengthen the administrative machinery of the state laws and to provide complete coverage of all indigent aged people within the state who qualify as citizens and residents. When approval of the state law has been granted, the federal treasury undertakes to bear half the expense of the pension system up to a maximum of $15 a month per person. In 1943–44 the federal government granted $340,776,000 to the states for old-age pensions; at the present time all but one of the states, as well as Alaska, Hawaii, and the District of Columbia, have approved pension schemes in operation. These schemes embody the features of the earlier state plans summarized in the preceding paragraph. In 1943–44 payments for old-age pensions totaled $720,000,000, of which the federal government contributed 47.6 per cent.

Old-age insurance: General. As distinct from old-age pensions, old-age insurance had been set up in many other countries before it was adopted in America. All told, thirty-two nations, including France, Germany, and Great Britain, were providing such insurance in 1938, either to all aged people compulsorily, as in Germany, or to those who voluntarily entered the scheme, as in Japan. In the United States until 1937 wage earners in private employment had available in old age only the pension plans described above, although compulsory contributory insurance had been established for

public employees in many cities and states and in certain services of the federal government.

United States Social Security Act. The old-age insurance provisions of the Social Security Act, in effect in 1937, provide a nation-wide plan for wage earners, with exception of certain categories of workers excluded by definition, under the direct administration of the federal government. In addition to government employees of all types, the following classes are excluded: agricultural labor, domestic servants, irregularly employed casual workers, the employees of religious, charitable, and educational institutions, and marine and railway workers. All other hired workers, regardless of the amount of wage or salary—the general manager as well as the manual laborer—are embraced in the system.

The details of the plan are complex, since they provide for a transition period during which the contributions to the reserves gradually increase until the scheme goes into full operation in 1949, but the essential features are these: Contributions are made in equal amounts by employer and worker in the form of a percentage tax on the payroll and on wages. In 1949 the employer is to be taxed 3 per cent of the wages of every insured employee, and the employee is to be taxed an equal amount on his wage income. The law provides for a gradual increase of the tax from 1 per cent in 1937 to the rate of 3 per cent in 1949, but thus far no increase has been made. In 1948 the payroll tax was still 1 per cent, though the original schedule stipulated a rate of 2½ per cent in that year. In practice the employer pays both halves of the contribution, deducting the worker's share from his pay envelope or salary check. The government makes no contribution toward the scheme. Funds raised in this way are carried in the old-age reserve account of the United States Treasury to be invested in 3 per cent federal securities.

Annuities are payable at the age of sixty-five, regardless of the financial circumstances of the individual, on condition that he ceases to draw wages at that time; as long as he continues in employment, the payment of the annuity is postponed. The amount of the annuity is a percentage of the worker's average monthly wage between 1937 and the date when he attains the age of sixty-five, with a minimum payment of $10 per month. An annuity one-half as large is payable to the worker's wife if she has reached the age of sixty-five, and a similar amount is payable on behalf of each

unmarried child under sixteen—or eighteen, if still in school. The total payment to a family on one worker's account cannot, however, exceed $85 per month.

Should the worker die before attaining the age of sixty-five, his dependents receive a lump sum payment of six times the monthly benefit which he would have received. In addition, a widow who has dependent children under eighteen, or who is herself over sixty-five and has not remarried, receives a monthly annuity three-quarters as large as that which her husband would have received. A dependent child under eighteen, or a dependent parent over sixty-five, receives an annuity one-half as large as the primary benefit.

Payment of benefits was begun in January, 1940, but the amounts paid are still relatively small because of the brief period during which benefit rights have accumulated.

The original plan, with its schedule of mounting payroll taxes, provided for the accumulation of a vast reserve to be invested in 3 per cent government bonds. It was calculated that by 1980 this reserve would total 49 billion dollars and that, from then on, the annual payments of benefits would be balanced by the annual receipts from payroll taxes plus income from the invested reserve.

This feature gave a false appearance of businesslike principle to the plan. Obviously a huge reserve invested in government bonds provides no source of financial support for the plan other than the tax system, upon which alone the government can rely for funds to pay interest on its own indebtedness. In other words, the benefits paid out each year must come from general taxes if not covered by the receipts from taxes on payrolls.

As a matter of fact, the payroll tax of 1 per cent has not yet (1948) been raised. Payments of benefits have been so much less than this rate has produced that the original tax has proved sufficient to provide an annual surplus to be held in reserve. At the end of March, 1948, the balance in the reserve fund was $9,632,000,000 and the fund was increasing at a rate of over $1,000,000,000 a year. Payments will of course begin to rise as the insured attained a higher average age. On the other hand, if the system is to be self-supporting, the payroll tax must eventually begin to increase, probably beyond what was originally planned.

Old-age insurance is too firmly established in modern countries and too strongly supported by economic and humanitarian con-

siderations to be open to serious challenge. The principle of requiring contributions from the beneficiary is sound. To require the employers to pay a share is probably defensible as an administrative measure, but it is somewhat misleading as to the final incidence of the cost. The employer will shift the tax, since it must operate as one of the regular costs of the business. This part of the contribution will eventually come to rest on the consumers of the products of industry, including of course the wage earners.

Industrial accidents: The problem. One of the principal costs of the modern system of industry is due to the temporary or permanent destruction of the productive power of the workers, through injuries resulting from the normal hazards of their occupations. Only within recent years has serious attempt been made to estimate the magnitude of this cost. The official statistics of work accidents in the United States are even now so unreliable and incomplete that any statement of the total loss can be no more than an estimate. In some states, no records at all are kept. Certain industries have accurate records of the fatalities resulting from the hazards of the occupation, though the less serious accidents which result in temporary disability are allowed to go unrecorded. Even on a conservative estimate, the loss can be shown to be surprisingly large. To give the result of only one study out of many, the National Safety Council found that in 1929 industrial accidents caused 20,000 deaths. The number of minor accidents has been variously estimated, usually in the neighborhood of 2,000,000 a year.

In recent years strenuous efforts have been made to reduce this waste. Factory acts and compensation laws have made it obligatory or profitable for the separate owners of business concerns to safeguard the lives and health of their employees. Groups of employers in the major lines of industry have coöperated through their trade associations to perfect devices for the prevention of accidents and to urge their adoption upon the more indifferent persons among their number. In the United States this effort has given rise to a widespread "safety first" movement, which enlists the services of experts and adopts many ingenious devices to make its program effective. Similar movements have been under way for decades in other countries.

The most significant result of these activities is their inability to solve the problem of occupational hazards. Indeed accidental deaths and injuries in industry have increased during this period,

both in absolute figures and in proportion to the numbers employed. They are caused by the nature of industry itself—the increasing mechanization of the productive process—to which the human mind and body seem unable to adjust themselves. As the basic cause is social, so is the ultimate loss—the temporary or permanent destruction of productive power.

Distribution of the loss: Individual loss. Though social loss from work accidents is inevitable, arrangements can be made which affect its distribution. Such arrangements do more than raise the question of justice. It has been found that different methods of apportioning the loss from work accidents have different effects upon the number of accidents and upon the magnitude of the social loss involved.

The burden falls in the first instance upon the injured workman and his family. Where the rate of accidents differs strikingly as between different occupations, this hazard will be one of the factors in the distribution of the labor supply. If the workers were entirely free to choose their lines of occupation, and if they made this choice on the basis of cool calculation of costs and income, the more hazardous forms of employment would be relatively undermanned, and the wage rate would be relatively higher than in occupations which were less hazardous. In this way compensation for the risk would be worked out by the forces of demand and supply, and the prices of the different kinds of goods would carry the cost on to the consumer. It is probable that a partial adjustment was made in this way before society undertook to distribute the risk by collective action.

But we have seen that the distribution of the labor supply is not controlled principally by rational choice, but rather by conditions of birth and tradition. The supply of labor in the hazardous occupations, as compared with other lines of work, is not materially limited by the risk of accident, and the wage rate, though it may contain a partial compensation for this factor of cost, does not suffice to lift the burden entirely from the workman.

Common law remedy. The workman has always been entitled to sue the employer for the recovery of his loss. In the courts the justice of his claim was tested by the principles of the common law defining the liability of employers. These common law principles were evolved at an earlier stage of industrial development to govern the simpler relations of master and servant, and their maladjust-

ment to the conditions of the modern world placed serious obstacles in the way of recovery by the workmen.

Stated briefly, the court attempted to discover who was to blame for the accident; if it could be shown that the fault was wholly the employer's, the workman was awarded damages; if it appeared that the workman himself or someone else shared the blame with the employer, no matter in how slight a degree, the case was dismissed without compensation. This was the essence of the common law *doctrine of negligence.* Without undertaking here to discuss all aspects of the common law remedy, it may be stated as a fact that the right to sue gave the workman very little protection.

Workmen's compensation laws. The public policy reflected in compensation legislation gives up all effort to assign the blame for accidents, recognizing that in the modern system of industry destruction of life and limb are inevitable normal costs not generally attributable to a particular person's fault. The financial burden involved in compensation for accident and death is accordingly placed solely on the employer. It is expected that this cost will become part of the enterpriser's expenses and enter into the price of the product, thus finding its way to the consumer, upon whom in the last analysis it is natural that the burden should fall.

One of the greatest virtues of the policy of holding the employer financially responsible for accident is that it operates to reduce the number of accidents. By making the hazards of occupation expensive to those who control industry, the compensation laws create an incentive in the group which has the chief power to prevent accidents. The "safety first" movement and similar campaigns for the reduction of accidents sprang up after society had determined to hold the employer responsible for the cost of injury received in the course of employment.

Modern compensation legislation began in 1884 in Germany, requiring employers to indemnify injured workmen or, in case of fatal accident, their families, and setting up an insurance system in which employers were obliged to insure this risk. In 1897 Great Britain adopted a comprehensive act, requiring indemnification at stated rates but leaving the question of insurance optional. France and Belgium have systems similar to that of Great Britain.

The first laws of the American states were declared unconstitutional, and it was not until 1909 that this policy received the approval of the courts. By 1945 all the states and territories, with the

single exception of Mississippi, had passed laws on the subject, as had the federal government for its own employees and the wage earners in the District of Columbia. Thirty-two of the laws, including that of the federal government, cover occupational disease as well as accident.

The essential feature of these laws is their requirement of compensation by the employer to the injured workman or, in cases of fatality, to his heirs. Aside from this common feature, the laws vary greatly in detail. Rates of compensation vary but are always a percentage of the injured worker's wages, paid in the form of life pension in cases of permanent injury, as a capital sum or an annuity to the heirs in case of death, and as a single payment for temporary disability. Many of the laws provide remuneration for hospital and medical costs as well.

In only twenty-three states does the compensation act require the employer to join any insurance system. The laws requiring insurance also vary; some provide a state insurance fund to which all employers must subscribe; others provide such a fund but leave the employer free to insure in private companies; and still others make no special provision for handling the insurance feature.

CHAPTER XXXIV

The American Organized Labor Movement

Structure of the labor movement: Conditioning factors. Unions of workers for collective action are to be found throughout the world wherever the modern system of industry has taken root. Appearing first in England, the birthplace of the factory system, these organizations have spread to all industrial countries and have flourished in rough correlation to the extent of industrialization. Today we may mark their beginnings in countries such as India and Egypt, which are only now taking over the productive methods of the more advanced nations.

Not only is there a close correspondence between the extent of union membership and the extent of industrial organization in a country, but within the most highly industrialized countries, where the movement is strong, unionism first appears and is strongest among those classes of workers who are directly affected by modern methods of production. In the United States there exist great numbers of wage earners on the farms, in retail trade, in domestic service, and in other positions of semidependence, among whom labor unions have as yet not gained a permanent foothold. In England, where the movement developed originally among industrial wage earners, unionism has recently spread to include agricultural and general laborers, retail clerks, teachers, and similar nonindustrial groups.

These facts indicate that the labor movement is an attempt on the part of great numbers of men to adjust themselves to certain conditions of life created by the modern system of industry. In the first place, under this system the worker is separated from ownership in the capital instruments and the land with which he works. The laborer is not an independent entrepreneur, and the mass of men in society find themselves in something like a uniform economic status; this tends to weld them gradually into a distinct social class.

Secondly, the workers are concentrated in close contact with each

other upon a limited geographical area. The discipline of the factory or mine, by accustoming the wage earners to a system of regimentation in which conscious coöperation is carried out under orders of management, provides conditions which are conducive to voluntary group conduct.

In the third place, there are present in the modern industrial system certain injurious economic forces which affect the workers in common and against which, as individuals, they are powerless to protect themselves. These forces have been sketched in preceding chapters; namely, unemployment, low wages, monotonous and hazardous forms of work, the labor of women and children—all problems whose solution is vital to the security and contentment of the mass of wage earners. The existence of these common problems supplies the motive for collective action among workers, as the regimentation of their productive life creates the condition which facilitates such action.

Types of American unions: Definition of labor union. We shall use the term *labor union* to apply only to independent organizations of workers having no organic affiliation with any employer or enterprise. In recent years recognition of the desirability of collective dealing has prompted many large employers in America to set up organizations among their own employees through which to carry out negotiations regarding wages and other conditions of work. These *company unions,* as they are called, present certain problems, which will be considered later in this chapter. They are not included in our definition of labor unions proper.

The type of organization adopted by wage earners when acting independently of their employers is adjusted to circumstances. The attempt is made to adapt the union structure to that feature of the industrial organization which exerts determining effect upon the conditions of the workers concerned. Since economic conditions are constantly changing, the structure of the labor movement is never exactly the same at different times or in different industries at the same time. It is possible, however, to distinguish two types of labor organization which contrast and even compete with each other in the general movement among American wage earners.

The trade union. The first of these is the *craft* or *trade union.* This type attempts to bring together all the workers who perform the same specific function and who are therefore likely to compete with each other. For example, bricklayers, carpenters, plasterers,

lathers, painters, though they all belong to the same major branch of industry, the building trades, are organized into separate unions in each locality. Similar craft divisions are to be found among the workers on the railroads; the locomotive engineers, the firemen, the trainmen, and the conductors are united into self-contained groups distinct from each other. In fact the craft union is firmly established as a type of organization in all branches of industry where skilled craftsmanship prevails. The local units of each craft have united into national organizations in response to forces tending to broaden the range of the labor market.

To bring about coöperation between the different unions, reliance is placed on the device of *federation.* These federations are both regional and functional. As an example of regional federation we find in each industrial centre the local craft unions federated into a city council for the purpose of solving problems of common interest in the locality. Similar federations within the area of each state attempt to bring about coherent action on legislative programs which fall within the power of the state legislature. At the top of the structure the American Federation of Labor draws the whole system together into a national grouping, whose chief component units are the national craft unions.

A functional federation takes the form of a loose affiliation of different crafts employed in the same general productive process. Thus the building trade workers, although allied primarily to their separate craft unions, are all drawn together into councils which attempt to promote intercraft coöperation on common problems. Similar functional federations are found in the printing trades, the metal trades, and on the railroads.

For many years this structure of independent craft unions, federated regionally and functionally for common action, has dominated the American labor movement. In 1944, out of 180 unions in existence, 130 were organizations embracing a single craft or narrow combination of closely related crafts. The American Federation of Labor has long advocated this form of organization as opposed to any other type, and its strong hold on the labor movement of this country gave it authority to speak for the organized labor movement as a whole. In recent years this unrivaled leadership has been challenged by the rapid growth under aggressive leadership of another form of unionism.

The industrial union. The second major type of union struc-

ture is the *industrial union.* This form attempts to obliterate craft and trade divisions among the workers and to make membership in the union coterminous with the limits of each separate industry. The factory system, with its machine technique and its refined division of labor, has tended to break down craftsmanship and undermine the position of the skilled workers. This process in the United States was accompanied during many years by a tide of immigration, which flooded the labor market with alien unskilled workers, thus tending to cause lower wages and presenting a threat to the higher standards of living of the native-born artisans.

When skilled and unskilled, native and alien, male and female struggle together to obtain employment, older lines of division between occupational groups tend to disappear, and the whole group of wage earners in each major line of industry tends to develop a consciousness of common interest. Organization along craft lines is out of place in this setting; adjustment to the situation calls for a type of organization which brings the workers together in a single union regardless of their diverse functions.

In 1944 there were 50 national industrial unions in the United States, some of them organizations with much power. The United Mine Workers of America is an example. In its local branches the U.M.W. brings together all types of labor working in the mines in a given locality—engineers, carpenters, draymen, unskilled laborers, as well as the actual miners of coal. These local units are merged into a national industrial union which aims to cover the entire mining industry of the country. The clothing trades are organized similarly, as are several others.

Some of the industrial organizations have joined the American Federation of Labor, despite its opposition to this type of unionism, in order to cooperate with the national craft unions in the solution of problems affecting the entire laboring population of the country.

Conflict of rival types of unions. The latent hostility between these contrasting types of union organization broke out into open conflict in 1935, when the Committee on Industrial Organization—a joint body of industrial union officials—launched a nationwide campaign to spread their type of unionism. The committee was formed primarily to attempt the organization of the large mass-production industries, such as steel, rubber, and automobiles, which were almost wholly nonunion.

It fell afoul of the A. F. of L. almost at once. It was outspoken in

its condemnation of the federation leaders for their apathy toward unionizing these industries. The C.I.O. also favored the industrial form of unionism, which brought it into conflict with certain craft unions, affiliated with the A. F. of L., which claimed jurisdiction over some of the occupations in these industries.

Despite the active opposition of the A. F. of L., the C.I.O. pressed its organization drive vigorously in 1936 and 1937, with dramatic results. Its aggressive tactics precipitated occurrences in the business world—the "sit-down strike," mass picketing, violent conflict with the police and the company guards—which aroused widespread national interest. Early in 1937 the C.I.O. won the recognition of the United States Steel Corporation, traditionally the leading anti-union concern of the country, in the form of a collective agreement covering all divisions of the corporation's enterprise. By the end of 1937, although it had lost bitter strikes with several large independent steel companies, agreements had been formed with some 450 companies in the metal industries, embracing a half-million workers.

During the period after 1935 certain federal laws facilitated the formation of unions in the unorganized industries and thus aided the C.I.O. in its campaign to unionize the great mass-production enterprises. The steel industry was thoroughly organized. In the automobile business the new industrial unions won recognition from all the large companies. Similarly successful were the efforts to organize the rubber and electrical industries. The outstanding accomplishment of this movement was to establish unionism for the first time among unskilled and semiskilled workers in the large-scale enterprises of the country. Beginning with an original membership of about one million, the C.I.O. now claims five million members.

The conflict between the A. F. of L. and the C.I.O. has focused attention on the relative merits of the two types of organization and promoted the belief that they cannot exist harmoniously side by side. As a matter of fact, the controversy does not turn entirely on questions of fundamental importance but is prompted in part by selfish considerations. The industrial unions, arising after craft unionism had become firmly established, have begun in certain industries to draw members away from the national craft unions, thus reducing their income from dues and the prestige of their officials.

The embittering effect of these selfish considerations clouds the basic issue, whether craft or industrial unionism is the better form to promote the interests of the wage earners. Whatever the outcome of the present struggle between them, it may be expected that in the long run that type of union will survive which is best adjusted to the prevailing economic conditions of the country.

Strength of unionism in America. The following table shows the growth of American unionism since 1900.

UNION MEMBERSHIP IN THE UNITED STATES

Year	A. F. of L. unions	All unions
1900	625,200	868,500
1910	1,630,300	2,184,200
1920	4,156,000	5,110,000
1930	2,961,096	4,100,000
1937	3,268,600	7,296,500
1940	4,247,400	8,500,000
1944	6,806,900	13,750,000

The figures for 1937 show the disruption of the union movement through the growth of industrial unionism under the C.I.O. Of the seven million members then outside the A. F. of L., the C.I.O. claimed five millions; the remainder were in unions, principally in the railroad industry, not affiliated with either federation. The figure for 1944 indicates the great expansion of union membership during the war due in part to a large and rapid increase in manufacturing employment.

By 1944 the industrial union had become the predominant form of labor organization. Excluding unions of government clerical employees and a few other unions not readily classifiable, 53 per cent of all union members in that year were in industrial unions, 36 per cent in craft unions, and 10 per cent in "mixed" unions.

Legal status of unions: General situation. The activities of trade unions have frequently been condemned by American courts, but at no time since their origin have unions as such been illegal in this country. The law of the land permits voluntary organization of individuals for the purpose of promoting their interests, and this permission applies to wage earners as to other interest groups. Accordingly, the right to organize and to act collectively has never been challenged by the legal system, provided the means used to effect organization and to carry out union policies have not been unlawful in themselves. Despite this legal immunity, the right of

organization has been fought by hostile employers with methods which have at times made organization virtually impossible.

National Labor Relations Act. With the adoption of the National Labor Relations Act of 1935, the federal government altered the attitude of American national law from merely permitting to positively approving and fostering union organization. This law listed a number of "unfair practices" of employers which were prohibited, including in the list many of the devices formerly used to prevent or retard union growth within the enterprise or to establish employer domination over the union when formed. Refusal to negotiate with the accredited representatives of an organization among the concern's employees was forbidden as an "unfair labor practice." The law created a permanent administrative agency, the National Labor Relations Board, made up ostensibly of members not partisan to either side of the industrial dispute, together with subordinate boards covering the principal industrial regions of the country.

Disregarding certain powers of the NLRB not pertinent to our immediate topic, it is to be noted that the board took aggressive action in favor of unionism. It employed the "unfair labor practices" sections of the law for this purpose, having been given broad authority to investigate, hold hearings, and make decisions of a judicial character whenever labor leaders in a given industry complained of "obstructive" tactics by the employer. In cases where the right of union officers to represent the employees was in dispute, or where there was disputed jurisdiction between independent unions and a company union, the board was empowered to take a vote of the wage earners, who thereby decided whether they wished to be organized and if so in what form of union. The outcome of such an election settled the question whether a given group of labor leaders was the accredited representative of the employees. If this question was settled in favor of the union, refusal to recognize the union's officers as a bargaining agency became an "unfair labor practice."

The act applied to industries, other than transportation, whose operations crossed state lines. This meant all large-scale industries of the country. Supplementing, however, the coverage of the act itself were the laws enacted by thirty-four states affecting labor relations in intrastate industry.

The statistics of the work of the NLRB reflects the gradual dis-

appearance of the activities of employers against which the act was directed. Between 1936 and June, 1944, the board held over 19,000 elections in which altogether more than 5,000,000 votes were cast. During the same period, it disposed of almost 35,000 cases involving unfair labor practices by employers. In 1936 four-fifths of the board's cases involved allegations of unfair labor practices. By 1944 the proportions were almost exactly reversed. Seventy-two per cent of the cases were representation cases and only 28 per cent were unfair practice cases.

However, it should not be assumed from these figures that employers had accepted the fairness and justice of the act. While proponents hailed the act as labor's "Bill of Rights," redressing the inequality of bargaining power of employees, the opponents viewed the law as unconstitutional and accordingly resisted stubbornly its provisions until the act was sustained by the Supreme Court in 1937. Thereafter the number of "unfair labor practice" cases decreased, but with only a temporary cessation during the war period in the barrage of criticism directed at the act. With the termination of the war in 1945, there ensued a succession of industry-wide strikes which not only set back postwar reconversion but also focused once again the attention of the public upon the possible threat of industrial strife to national health, safety, and prosperity.

Few of the provisions of the National Labor Relations Act escaped attack. It was charged that the act was unjustly discriminatory, since it outlawed only "unfair labor practices" by employers and imposed only upon employers the obligation to bargain collectively. Employers objected to the exercise by the NLRB of both administrative and judical powers. They sought the same rights accorded labor organizations to petition the board for employee elections and demanded that there be imposed upon labor organizations the same restraints upon coercion of employees as applied to them. Those provisions and practices under the act which gave governmental encouragement and support to the closed shop were condemned.

Labor Management Relations Act. Early in the 1947 session of the Congress, the Senate Republican majority recognized these criticisms and announced as one of the objectives of the session the enactment of legislation "to achieve a better balance between legal rights and responsibilities of employers and employees and unions and by other means to lessen the number of strikes that interfere

with commerce . . . [and] to provide a practical and sound solution for the industry-wide shut-downs which have a paralyzing effect on our economy." On June 23, 1947, there was enacted into law the Labor Management Relations Act of 1947, popularly known as the Taft-Hartley act.

This act amends the National Labor Relations Act of 1935 by reconstructing the NLRB, by broadening the legal rights of employers, and by increasing the liability and responsibility of labor organizations. The membership of the NLRB is increased from three to five, and most of its administrative duties are transferred to its general counsel, so that as now constituted the board is restricted primarily to the judicial functions of the act.

To the former list of "unfair labor practices" of employers is added a list of "unfair labor practices" of labor organizations. These include coercion of employees in their rights relating to collective bargaining, discrimination against employees under union shop contracts for any reasons other than the payment of dues and initiation fees, refusal to bargain collectively with employers, engaging in certain types of strikes and boycotts, exacting excessive and discriminatory fees under union shop agreements, and engaging in the restrictive practice known as "featherbedding." These and other restrictions imposed upon the activities of labor unions will be discussed in later sections of this chapter.

Bearing more directly, however, on the legal status of unions as such are the provisions of the act imposing new duties upon unions as conditions of being recognized as representatives of employees for collective bargaining. To qualify for such rights under the act, unions are required to file with the Secretary of Labor annually reports relating to their organization, procedures, and financing, and to submit to the NLRB affidavits of their principal officers swearing that they are not members of or connected with the Communist Party or any other organization advocating the overthrow of the United States Government by other than constitutional means.

Company unions. The unionism we have discussed is a class movement in opposition to the managerial and ownership factions in society. It views the interests of these two classes as antagonistic and attempts to cut the industrial system horizontally through the boundaries of the separate business enterprises.

Another type of organization has developed in recent years, called the "company union." This form organizes both managers and

workers into a single unit within each separate enterprise, on the assumption that their interests are mutual and that both groups will benefit from any device which promotes harmony and efficiency. The structural details of these company unions vary among different enterprises, but they all have as their essential feature a system of councils jointly representative of labor and management. These councils discuss and exercise some measure of control over business policies which most closely affect the workers. The matters thus dealt with include wage rates, methods of wage payment, discharge of workers, and the like.

The spread of company unions was stimulated after the First World War by the aggressive policies of the independent unions, which caused many large employers who believed in the principle of collective dealing to organize their own employees within the limits of their separate enterprises. A clause of the National Industrial Recovery Act of 1933 making collective dealing obligatory further stimulated this growth.

A bitter struggle between the independent unions and the company unions was a striking feature of industrial relations in the United States until 1936. The independent unions believed that the rival form was a creation of hostile employers whose real purpose was to destroy the union movement and weaken the bargaining power of labor. The company union was always initiated by the employer, a fact which made labor leaders suspicious of its purposes. Its effect was to tie the workers up with their respective enterprises, separating them from their fellows in other business concerns and other localities.

The National Labor Relations Board has assisted independent unionism in its struggle with the company unions. Among the "unfair practices" prohibited by the National Labor Relations Act, and continued in the Labor Management Relations Act of 1947, is any attempt on the part of the employer to dominate any labor organization in his enterprise, even if the organization is a company union. Since most of the existing company unions were created by the employers, the board has ruled that many of them are incompetent under the law and has called elections in the enterprises in question to determine whether the workers wished to be represented by the independent unions. These activities of the National Labor Relations Board hastened the decline of company unions after 1936.

The program of American labor unionism: Political activity. In one regard the American labor movement has not followed the lines of development marked out by the history of the older movements of European countries. During the past three decades, in which period the unions in the United States have made their most substantial and continuous gains in strength, they have abstained from organizing an independent labor party. Most of the larger national craft unions, as well as the A. F. of L., have incorporated in their constitutions provisions which forbid their members from attempting to resolve the organization into a political party. With the recent growth of industrial unionism there has been some departure from this practice, the C.I.O. having entered the political arena in certain important industrial areas, sometimes with its own candidates for office.

Nevertheless it may still be said that on the whole American unions conceive their problem to be primarily economic rather than political; their objective is to enhance their bargaining power as against the employers and not to capture the power of the state in the effort to put through far-reaching reforms. Accordingly their members are urged to retain their independence of choice of party membership, the official action of the unions in the way of influencing this choice consisting only in their endorsement of those candidates presented by the major parties who have shown themselves "friendly to the interests of labor."

Recently the C.I.O. unions have applied this policy with greater vigor and effectiveness through a central Political Action Committee, which exerts pressure on the nominating conventions of the parties and takes an active part in election contests. In addition to the nonpartisan endorsement of candidates, the unions exert political influence by maintaining strong lobbies at the state and federal capitols. However, the Taft-Hartley act has cast doubt upon the legality of some phases of these activities. This act makes it unlawful for any labor organization to make a contribution or expenditure in connection with any national election.

This policy of indirect rather than direct political action is a practical adaptation to peculiar American conditions. Several early attempts to organize labor in America were destroyed by ill-advised campaigns in politics, and the unions have learned to fear the divisive influence of partisan politics upon their membership.

The American governmental structure is such as to discourage in-

dependent activity on the part of any economic group. The division of sovereignty between national and state governments with the important power over social legislation distributed among the forty-eight different commonwealths, the separation of the governmental machine into semi-independent executive, legislative, and judicial units, the power of a Supreme Court independent of the legislature to pass upon the constitutionality of state and federal statutes, these and other characteristics of the system multiply the obstacles in the way of effective political action on the part of any minority economic group operating independently.

Moreover the predominantly *laissez faire* character of governmental policy has tended, at least until the advent of the "New Deal," to minimize the importance of legislation as a means toward the accomplishment of economic ends and has encouraged workers and employers alike to rely upon their own bargaining power.

A few of the strong unions have consistently asserted the necessity of carrying on independent political action along lines similar to those followed by the labor movement of England. It is noteworthy that the unions which sponsor the formation of an independent labor party in this country are almost wholly unions of the industrial type, whose broad inclusive basis of membership compels them to adopt policies of general social import. The attitude of the C.I.O. has been noted. The tendencies which further this form of unionism therefore tend also to promote the political activity of organized labor.

Other tendencies may be observed working in the same direction. In all industrial countries governments are becoming an increasingly important force in economic affairs, regulating business activities, establishing standards of economic conduct, restricting individual initiative along many lines, entering the industrial world as owners of productive enterprises and as employers of labor. These expanding economic functions of government may attain such a degree of development that organized labor can no longer accomplish its aims by economic action unsupported by its organized political power, in which case we may expect the unions to organize for political action.

Social features and benefit systems of unions. Labor organizations perform many incidental services for their members in addition to the functions which are their essential reason for existence. Most of them act as fraternal associations, promoting social activities

among those who belong to them. The stronger unions provide insurance and charitable relief for the victims of unemployment, sickness, or accident. These benefit systems in the aggregate reach multitudes of people and provide an important element of security in the lives of the wage-earning population. In 1928 it was found that 78 of the unions in the United States had such systems in operation. In 1944 the unions affiliated with the American Federation of Labor paid out some $19,000,000 to their members.

The interest of unions in benefit systems received special emphasis when, in negotiating for a new contract in 1946, the United Mine Workers made a demand for a welfare fund to be financed by a royalty on each ton of bituminous coal mined. When these negotiations failed and a strike ensued, the United States Government seized the mines and concluded arrangements with the union for their operation, including the establishment of the welfare fund. The Labor Management Relations Act legalizes agreements providing for employer contributions to union welfare funds, when established solely for medical or hospital care, retirement and survivors' pensions, compensation for injuries or illness resulting from occupational activity, unemployment benefits, and insurance covering life, disability and sickness, and accident, but with restrictions designed to safeguard the use and custody of such funds.

Important as are these activities, however, they do not constitute the distinctive economic function of labor organization. This consists in the attempt of the union to raise wages, shorten hours, and improve working conditions through the use of collective bargaining. There are some unions—the revolutionary wing of the movement—which look beyond these immediate material gains to an eventual overthrow of the capitalist system of industry, but in America these unions comprise but a small minority of all organized labor, and they too practice collective bargaining as their normal day-by-day function.

Collective bargaining: Purposes. The essential feature of collective bargaining is the requirement that the employer hire an entire labor group as a unit, the terms of employment being settled by negotiation with representatives of the labor force. From the worker's point of view the merit of this method is derived from the bargaining weakness of the individual laborer in the modern system of industry, his ignorance of the prevailing wage or of his real worth to the employer, his timidity in asserting his claims, and his

urgent need of a job, which deters him from withholding his labor from the market until better terms are offered.

Our study of the law of wages has made it clear that the actual wage in a given case need not be equal to the productive worth of the worker, and it can hardly be expected to settle at this point if the two parties to the wage contract are very unequal in bargaining power. Low wages are not always a sign of exploitation, nor are high wages always proof that the worker is receiving his economic deserts. The test in each case is whether the rate actually measures the marginal productivity of the labor group in question. When the wage is below this point, whether it be high or low in amount, it can be raised by a change of bargaining methods which substitutes the collective for the individual contract.

This principle also marks out the limits within which permanent improvement can be obtained merely by more aggressive bargaining; the wage can be advanced beyond the point of marginal productivity only at the cost of throwing some of of the laborers out of work. The history of trade unionism affords many instances in which collective bargaining has permanently increased the real wages of the organized workers.

It would underrate the importance of collective bargaining to the wage earners to assume that it is concerned solely with the question of wages. As used by modern unions, this device is employed to give the worker greater security in the control of his job. The collective agreement customarily prescribes methods of hiring, assignment of work, transfer and promotion, layoffs, and discharge. It obliges management to share with the union officials decisions on these important questions, as well as on the operation of day-to-day discipline in the plant. These features of the collective agreement are no less valuable to the employee than are those which have to do with rates of pay and hours of work.

Benefits of collective bargaining to employers. Collective bargaining, however, need not be of advantage to the laborers alone; it results in certain indirect benefits to the employer, which enhance its value in the industrial system. The mechanism through which this type of bargaining is carried out may be turned to account in drawing workers and employers together into closer contact on a basis of mutual tolerance and good feeling. Many of the wastes in modern industry and much of the friction and bitterness

engendered in the field of industrial relations flow from the absence of such contact.

Organization of the workers for collective bargaining and the establishment of a permanent structure for negotiation with the employers may affect a real improvement in the situation. The workers are supplied with a mechanism for voicing their grievances and are assured of a respectful hearing. Discipline in the shop is more effective when it bears on observance of the terms of a contract in the making of which the men have had equal voice. When subordinate officials are restrained from harsh and overbearing tactics, the improved treatment of the workers increases their contentment and their loyalty.

Above all, labor is given a recognized share in the control of those features of industrial policy which most vitally affect their welfare, and this also tends to give them a more cordial attitude toward the enterprise. It goes far toward supplying a substitute for that personal interest in the product and the fortunes of the concern which existed in the handicraft stage of economic evolution but was weakened, with far-reaching results, by the growth of the factory system. These benefits from organized collective bargaining are now recognized by practically all modern employers whose enterprises have grown to any size.

Conflict over method in collective bargaining. But though the soundness of the principle of collective bargaining is generally conceded, there is no such agreement regarding the method of applying this principle. Organized labor insists that the workmen in each of the different enterprises must be represented in their bargaining by the regular officials of the union. These men are not on the payroll of any employer with whom they negotiate wage agreements but are professional bargainers representing all the members of their union within a given geographical area. Recognition of the union binds the employer to adopt this method of bargaining. On this point the issue is joined between the independent unions and employers adhering to the company union principle. These employers, though they are willing to deal with their workers collectively, insist that the workers' representatives must themselves be employees of the enterprise.

Strong arguments are presented in defense of each of these opposing systems of bargaining. Employers who refuse to recognize the union insist that the "outside" negotiator is ignorant of the

peculiar circumstances of the business and therefore unable to discuss terms adjusted to specific needs of that business. Moreover these professional union officials are frequently so aggressive and domineering that peaceful compromise is impossible.

On the other hand, leaders of organized labor stress these weaknesses of the company union as a device for protecting the interests of the workers: (1) the individuals chosen to represent the workers lack courage because their jobs depend on the good will of the employer, (2) these representatives, being ordinary wage earners continuously employed at their jobs, can acquire no knowledge of general trade conditions and no expertness in negotiation, (3) final control always rests with the management, which may veto the decision of the council and abolish the whole system at will, (4) this method of organization binds the wage earners to their individual enterprise, subjecting them to its hazards and depriving them of the collective support of other workers in case they are driven to strike in self-defense.

The controversy over the method of bargaining rarely discloses the real motives of either side. To organized labor, recognition concerns not only a process of bargaining but the very life of the union itself, and it is for this that labor leaders are fighting when they oppose the company union. On the other hand, the employer's resistance to the union's method of bargaining is really inspired by fear of certain practices to which organized labor may resort when sufficiently entrenched in the industry. In particular it is the closed shop with all its possible abuses that moves employers to refuse to recognize the union.

Regulation of collective bargaining by the United States Government. The Labor Management Relations Act of 1947, as did its predecessor, the National Labor Relations Act of 1935, disposed legally of the controversy over collective bargaining for industries affecting interstate commerce, by declaring the right of employees to self-organization and "to bargain collectively through representatives of their own choosing," and by prescribing certain of the conditions governing the collective bargaining process. Under the 1947 act, an obligation is placed upon both employers and unions to bargain collectively in good faith.

The employer must bargain with that union as the exclusive bargaining agent of all employees in any particular unit of his establishment which by a representation election conducted by the

NLRB or by other means has been selected by a majority of such employees to be their representative. Neither the employer nor the union may coerce employees to join or not to join a union by discrimination in connection with their employment or conditions of work.

The freedom of collective bargaining is further safeguarded by making it unlawful for any employer to pay money or any other thing of value to a representative of his employees except for check-off of dues and contributions to welfare funds under rigidly prescribed conditions and for a few other minor purposes. A sixty-day notice is required of either party for the termination or modification of any collective bargaining agreement, during which period the contract continues in full force and effect and any strike or lockout is illegal.

The closed shop. Strictly defined, a closed shop is one in which none but union men will be employed to perform those tasks for which the labor supply has been unionized. Its opposite, the open shop, is one in which there is no discrimination in offering employment as between union and nonunion men. Both of these terms are often ambiguously used. For example, some employers who declare their adherence to the open shop principle announce that they will not hire union men at all. This policy obviously does not lead to a real open shop.

Moreover the closed shop has been confused with the *closed union* in much discussion of its merits and defects. A closed union is one which refuses admission to properly qualified workers. The effects of the closed shop are obviously quite different according as this privilege is enjoyed by a closed union or a union open to all comers, and it is a mistake to assume that the closed shop and closed union are necessarily found conjoined.

The closed shop is but an extension of the union program of collective bargaining, which is defective unless the contract is binding on the entire labor force. When union and nonunion men are working side by side it is difficult to enforce the standards set in the collective agreement because of the employer's power to accord nonunion men different treatment and to show them preference when expanding or decreasing the work force.

Aside from this general benefit, the closed shop is of advantage to the union. It enhances the worker's inducement to join the union and reduces the difficulty of expanding the field covered by or-

ganized labor. It strengthens the disciplinary powers of the union officials; the recalcitrant member expelled from the union suffers unemployment. This in turn facilitates the collection of union dues and makes for faithful observance of union rules. It is an absolute guaranty against the victimization of spokesmen for the workers when defending the interests of their constituents. It strengthens the hand of the union when, in the last resort, it orders a strike. In fact the closed shop makes the strike so formidable that the threat of the strike often suffices to win the point for which the union has been contending.

On the face of the matter the closed shop may be an advantage also to employers by giving the union officials power to enforce faithful observance of the collective agreement and to guarantee a standard of performance. On the other hand, there is obvious danger that these officials will use their power to encourage time-wasting practices of the workers, to prevent adoption of labor-saving machinery, to oppose the efforts of management to improve labor efficiency, and the like.

The closed shop is a form of monopoly. Strongly entrenched unions enjoying this privilege may restrict the admission of other qualified workers to their organization, thus increasing the scarcity of their labor and raising their wages at the expense of excluded workers, and eventually to the damage of society as a whole. They may also restrict output while on the job and in this way reap gains from higher prices. These are not merely imaginary dangers; many closed shop unions have put these and similar monopolistic devices into practice.

Collective bargaining and the closed shop are devices useful to labor in that they enable the workers to obtain the full reward for their services as set by economic law. They may go along with an open union policy and the absence of restrictive or monopolistic devices in the program of organized labor. But unless supported by such devices, their effectiveness to improve the economic condition of labor is limited.

The real income of the different groups of workers is controlled fundamentally by the numbers in these groups in relation to the demand for their services. An overcrowded group may make full use of collective bargaining, standardization, and closed shop and yet, if all its members are to find employment, they must be content with a small income. Recognizing this fact, the strongly entrenched

unions are at times prone to adopt various monopolistic practices.

This is not to say of course that the powers given union leaders by the closed shop are invariably abused. The basic principle is that monopoly, whether of employers or of laborers, generally operates against the optimum production and distribution of the goods that satisfy human wants.

In industries subject to its provisions, the Taft-Hartley act outlaws the closed shop and authorizes the *union shop* only where the union has been designated by the employees as their exclusive bargaining agent and has been certified by the NLRB as having authority to make such an agreement. Under the union shop, an employer may hire nonunion workers but may retain them in employment only if within a specified period thereafter they join the union. *Maintenance of membership* clauses, under which the employer must discharge an employee who drops out of the union at any time other than that allowed by the agreement, are also permitted by the act, but only under substantially the same conditions as govern the union shop.

Restrictive practices: Membership. Some unions adopt the practice of artificially restricting their membership. This is done, not by direct refusal to admit qualified workmen, but in certain indirect ways which reach the same result; for example by setting up prohibitive initiation fees or requiring an unreasonable period of apprenticeship.

No union would dare to adopt such a policy until it had obtained a sufficient degree of competitive power. The closed union cannot exist under open shop conditions, for the reason that the members of the union would then be compelled to compete for employment in the same industries with the excluded workmen, whose competition would be the more destructive of union standards because they were unorganized. Also the right of collective bargaining through union officials must have been won before there can be any incentive to close the union.

Collective bargaining and the closed shop take on a sinister aspect when they are used as the indispensable first steps toward the creation of a monopoly of labor. Though by no means a general practice, this is done by some of the strong unions. By applying the policy of restricting its membership, an organized group makes itself a scarce factor in industry, raising the value of its marginal man's product and hence the wages of all its members. These

gains are made at the expense, in the first instance, of the excluded workmen who are forced to render service of a type which is more plentiful and of lower value than they are qualified to perform. In the long run all members of society share the loss through the decline in the social income which must result from employing a part of the labor force at relatively unproductive tasks.

Restriction of output. A second type of restrictive practice consists in a variety of devices for reducing the normal output of union members. Most wage earners are disposed, under certain conditions, to restrict their output and adopt "make-work" policies. In large part this is a reaction against the menace of unemployment; in part also it is a deliberate attempt to enhance the value of the product by increasing its scarcity. Restriction of output is not a unique policy of organized, as distinct from unorganized, labor, though it is true that the trade union is in position to apply the policy with greater effect.

Whether labor can or can not restrict output depends in the main on the technical nature of the productive process. Conditions are most favorable when a craftsman produces a finished product. Thus the classic examples of restriction of output have occurred among craft groups, bricklayers, stonecutters, glassblowers, etc. An industrial union has much less chance to pursue such a policy, because the work is frequently machine-paced, and any attempt to hold down output must take the form of control of machine speeds.

The strike: Definition. The last resort of organized labor when the bargaining system breaks down is to settle the issue by open warfare. Strictly defined, the *strike* is collective quitting of work by laborers who do not intend to surrender the job permanently. Correlative with the strike is the employer's weapon, the *lockout.*

Spontaneous quitting of work has always been the practice of wage earners when other methods fail to win for them acceptable terms of employment. When however an organized group of workers strikes, it does so with the intent of retaining control of the vacant jobs until an agreement is reached. This is an important element in every strike situation, though it is not always understood by the general public. The result of this attitude of organized labor is violent opposition to any workers who attempt to take the jobs while the strike is in progress. Such people are viewed as "scabs" and "job stealers," against whom strong methods are considered legitimate.

All strikes involve loss to the workers; hence the weapon is not used thoughtlessly but is generally resorted to only after deliberation and in defense of what the workers conceive to be a vital element in their standard. There are exceptions to this rule, particularly in America, where racketeering has frequently contaminated the labor movement; corrupt union officials will call strikes in their own personal interests or to victimize an enemy among employers or rival labor leaders.

Although workers when well organized and well led use the strike weapon with caution, still the menace of the strike has increased in direct ratio with the spread of unionism. Strikes occur much more frequently today than when the union movement was in its infancy. They take place on a wider scale and are more protracted and more bitterly waged, and each strike is more of a national calamity, embracing a wider section of the industrial area and paralyzing the sources of supply for goods essential to public health and safety.

Statistics of strikes in the United States. The magnitude of the strike problem is shown by the table below. Strikes tend to increase in number and magnitude during prosperous years and to decline during depression. Periods of numerous strikes are also correlated with organization campaigns of the unions.

TREND OF STRIKES IN THE UNITED STATES, 1919–1937[1]

Years	Strikes	Strikers	Average days lost per year
1919	3,630	4,160,000	140,000,000
1921	2,890	1,110,000	56,000,000
1932	840	320,000	5,100,000
1937	4,740	1,861,000	28,400,000
1941	4,288	2,363,000	23,000,000
1942	2,968	840,000	4,100,000
1944	4,956	2,116,000	8,700,000
1945 (est.)	4,500	3,285,000	27,500,000

These relations are shown by the figures given for individual years in the table. The year 1919 was one of flourishing business activity, marked by an organization drive upon certain large-scale nonunion industries; 1932, near the trough of the depression, was a year of comparative industrial peace; 1937 shows the effects of

[1] Figures through 1932 adapted from Daugherty, C. R., *Labor Problems in American Industry*, Revised edition, Boston: Houghton, Mifflin Company, 1938, p. 287. Later figures from *Monthly Labor Review*, May, 1945, p. 958, and January 1946, p. 85.

the C.I.O. campaign during the recovery period immediately preceding.

At the outbreak of the Second World War most unions laid the strike weapon aside in consideration of the government's support of collective bargaining and unionization. The result is shown in the decline in strikes in 1942. Before the war's end, however, strikes had begun to increase, and at the cessation of hostilities there was an immediate outburst of widespread strikes, whose effect is shown in the 1945 figures for number of strikers and number of days lost. This condition carried over into 1946.

The interrelation of strikes with the attempts of unions to expand their membership is of particular importance in American labor history. When the attempt of the union to organize the workers in nonunion industries is fought by the employers, a strike is the inevitable result. The great steel strike of 1919, the shopcrafts strike on the railways in the early 1920's, the steel and motor industry strikes of 1936 and 1937 are examples. Certain aspects of the industrial conflict in this country—destruction of property, violence to persons—are also identified with these periods. The nationwide strikes of 1945 and 1946, however, did not result from employer opposition to unionism but from disputes between workers and employers over wages, hours, and conditions of employment.

Types of strikes. Most strikes are what are called *primary strikes; i.e.,* called by a group of workers because of a disagreement with their own employer over their own terms of employment. Sometimes, however, workers strike who have no grievances of their own and who ask for themselves nothing in the way of better terms. These are *sympathetic strikes,* called by one group of workers for the sole purpose of aiding another group in a different business concern.

Sympathetic strikes are usually limited to two types of situations. In the first case, the employer involved in the primary strike may be carrying on his industry by the use of strike breakers, and union men everywhere are anxious to prevent him from marketing his goods. In this case the sympathetic strike is called to coerce another employer, sometimes in a distant city, into withdrawing his patronage from the concern in question. In the second case, if the primary strike is a fight to preserve the union from destruction, organized labor of the locality in quite unrelated industries or the

other unionized crafts in the same industry may strike in support of a principle considered vital to all unionized workers.

A sympathetic strike may reach such dimensions within a localized area as to call out all unionized workers in that locality. There have been such *general strikes* in the United States, each of them limited to a single city and effective for a short time in bringing the entire economic life of the city to a standstill.

It is not difficult for labor leaders to construct a theoretical defense of the sympathetic strike or even of the localized general strike. The interdependence of industry has developed to such a point that a strike in one business unit may be very seriously affected by what is done in other business units in distant localities. The loss of a strike called in defense of some vital principle of unionism may seriously weaken the position of unionized workers elsewhere. These considerations are especially persuasive when employers are united in nationwide associations in opposition to the spread of unionism.

However, while subscribing to these ideas in the abstract, the officers of the American national unions do not favor the use of the strike in these secondary forms, for the very practical reason that sympathetic strikes are rarely successful and general strikes never so. The public at large reacts to a general strike as to an insufferable attack upon its security and safety and turns out *en masse* to crush the revolt. Similarly the employer suffering from a sympathetic strike has little difficulty in convincing the public that the strikers, having no grievances against his labor policy, are acting in an indefensible manner. These strikes do occur in America, but they are almost invariably called by local union officials against the advice of their national officers.

Violence in strikes. Every strike creates a situation in which violence to persons and property is very likely to develop, and many strikes have degenerated into a state of disorder closely bordering on civil war. It is noteworthy, however, that violence is normally associated either with an organization campaign in nonunion industries or with strikes called to prevent the extinction of a union when an employer who has formerly dealt with it has refused to continue recognition.

In such cases the union is fighting for its life, and the situation calls to the front the aggressive, bellicose type of leader rather than the statesman. There is an uncompromising attitude on both sides,

which creates a spirit more akin to war than to business negotiation. The widespread sit-down strikes of 1936 and 1937, which were tantamount to forcible seizure of the company's property, were a novel development of the war tactics of newly formed unions in such a situation. Most of the earlier instances of bloody and destructive conflict in American labor history—such as the Homestead Strike in 1892—developed out of similar situations.

When violence occurs in strikes called by established and recognized unions in the course of bargaining over the terms of employment, the cause is usually the presence on the scene of conflict of nonunion strike breakers or the police force or militia. As pointed out in another place the free lance worker who accepts employment in an industry during a strike is viewed by the strikers as a "job stealer," against whom methods of coercion may fairly be used when milder persuasive devices fail. As for the presence of the police or militia, this usually results from an appeal that property rights and the right of the nonstriker to work are entitled to protection. Since these appeals always come from the employers and property holders, and never from the strikers and wage earners, the appearance of the militia on the scene of action is taken by the strikers as evidence that force is to be used to aid the employer to win the strike.

The practice of mass picketing developed in recent years by the C.I.O. unions has injected into the strike situation a new danger of violence. These tactics are adopted partly to strengthen the morale of the strikers, partly to exclude from the plant supervisory and managerial staffs who are not unionized. Mass picketing has resulted in violence in certain recent strikes which were called by well-established unions.

Legality of the strike. In the United States there is no single law governing the activities of the unions, since this is a matter lying within the province of the states. As regards strikes, state laws are so frequently changed and the opinions of state courts are so at variance with each other that it is impossible to phrase a general rule regarding the legality of the strike in this country.

The only approach to a general principle is this: all strikes are legal, provided their object is clearly and solely to improve the economic condition of the strikers; *e.g.*, by raising their wages, shortening their hours, or bettering their conditions of work. Unless the strikers can prove that their whole intent is to benefit themselves, not to work injury upon others, this rule of law does not justify

their conduct. Under the American law all combinations are criminal conspiracies when their purpose is to destroy property or to interfere with the rights of others. As a general thing the local primary strike usually survives the test of motive, but strikes on a wide scale, and especially the sympathetic and the general strike, are likely to be prohibited.

When strikes are held to be criminal conspiracies, the courts will attempt to prevent them or bring them to an end by the use of the injunction. This legal device has broken many strikes, but since it leaves the causes of the controversy unchanged, it offers no permanent solution of the problem but only leaves the workers with an increased feeling of bitterness. Certain strike weapons are sometimes prohibited even in legal strikes—for example, picketing and the "sit-down"—as working injury on innocent people not party to the dispute or openly violating the property rights of the employer.

The Labor Management Relations Act declares unlawful the sympathetic strike, as well as any strike by employees of the United States Government. It also provides for special governmental intervention in strikes or lockouts affecting an entire industry, or a substantial portion thereof, in interstate commerce and threatening national health and safety.

Public intervention in strikes: The problem. The public is always the third party to every strike, if only in the role of victim. As industrial disputes have increased in number and scope, the public has been compelled in its own defense to seek ways and means to bring order into this vital area of its life. In the absence of alternative methods for the adjustment of the workers' grievances, it has been found impossible to stop strikes by the simple expedient of prohibiting them. Modern efforts to solve the problem therefore consist in setting up machinery for the adjustment of the dispute before it reaches the stage of an open break.

Conciliation: Scope. The least formal of the measures used to effect a speedy settlement of a trade dispute is *conciliation.* Strictly defined, conciliation denotes a method of settlement in which the employers and the workers, or representatives of these two groups, voluntarily meet together for discussion and agreement without aid or compulsion from the outside.

Conciliation implies a high degree of organization on both sides and a habit of collective negotiation. It is a natural outgrowth of

the development of trade-unionism, on the one hand, and employers' associations, on the other.

The method is best used for interpreting the meaning of doubtful points in an existing trade agreement or of working out a peaceful compromise between conflicting demands of an employers' association and a trade union, to be embodied in a future agreement. Such differences as these usually relate to specific practical matters of detail, on which compromise is both possible and beneficial to both parties. In the absence of machinery which begins almost automatically to function for the settlement of these differences, the alternative is an open break and a trial of force which causes an immediate loss to all concerned and leaves an aftermath of bitterness and resentment.

The effectiveness of voluntary conciliation is not fully appreciated by the public, because in the nature of the case this process, if successful, forestalls a disturbance which might have attracted public attention. As a matter of fact, conciliation has a long history of orderly and successful operation in many highly organized industries.

Questions of general principle, such as the right to organize, the prerogatives of union officials, etc., cannot be settled by conciliation. Agreements to bring such matters within the range of conciliation almost invariably cause a breakdown of the whole machinery of friendly negotiation.

Furthermore experience has proved that it is necessary for the conciliation board to reach a unanimous decision on all matters referred to it. The board is made up of equal representatives of workers and employers; its proceedings are informal; its decisions cannot be enforced against a recalcitrant party. Unless decisions are unanimous, the constituents of those representatives who have voiced dissent are very likely to repudiate the verdict. A strike or lockout ensues, which brings the whole process of conciliation into disrepute.

Government encouragement of conciliation. Public intervention occurs when the government requires industry to organize conciliation machinery in areas where collective bargaining, the trade agreement, and joint conference have not been established voluntarily. Some countries have set up the joint council system of conciliation throughout the industrial field as a whole, but the adoption of this plan has not been possible in the United States,

because of the imperfect organization of the wage earners and the refusal of many large employers to deal with independent unions.

Attempts of the federal government to encourage conciliation have frequently broken down in the face of the employer's refusal to accept the principle of collective bargaining through union officials. This was the fate of the conciliation machinery set up by the transportation act of 1920 and the Watson-Parker act of 1926. The national and regional labor boards set up in 1933 and 1934 as a phase of the NRA were designed as machinery for mediation and arbitration, but they were based on the expectation that a wide structure of conciliation councils would be called into existence by the legal requirement of collective bargaining. In this case too the substructure of conciliation was disrupted in certain industries by the refusal of employers to recognize the independent unions.

Governmental attempts to promote conciliation in America will continue to cause more industrial disturbance than industrial peace as long as the company union and the independent labor organization are in conflict with each other. At the present time, as noted in a preceding section, the boards set up by the National Labor Relations Act of 1935 and by subsequent state laws promoted the spread of independent unionism at the expense of the company unions and laid broader foundations for conciliation.

Mediation: General nature. Mediation is the first step in the direction of acknowledging the right of the public to a voice in the affairs of employer and employee. It involves a friendly and informal effort, by some individual or agency not party to the dispute, to bring the disputants together for the purpose of working out a peaceful settlement. The function of the mediator is diplomatic rather than judicial.

That informal, noncoercive intervention of this sort is so frequently successful may be attributed to human nature. The mediator is usually a man of prominence in the social or political world, who comes forward at the critical moment to volunteer his services. In America mayors of cities, governors of states, the Secretary of Labor, and even the President have performed this function with success. The eminence of such a mediator flatters the disputants with a sense of their own importance and inclines them to make concessions as a personal favor to him which they would refuse to make to an opponent.

Official mediation in the United States. Many countries have provided regular machinery for mediation in the form of standing boards of public officials. In the United States the Erdman act of 1898 provided that the Chairman of the Interstate Commerce Commission and the Commissioner of Labor should offer their services as mediators in disputes on the railways. Power to act as arbitrators was also given to these officials. Between 1898 and 1912, forty-eight applications for mediation and arbitration were made to this board, the majority of the cases being settled by mediation.

When the Newlands act superseded the Erdman act in 1913, provision was made for a Commission of Mediation and Arbitration to be appointed by the President for a term of seven years. Of the seventy-one cases in which the services of this commission were employed between 1913 and 1917, most were settled by mediation.

The law of 1913 which created the Department of Labor set up a mediation service under the Secretary of Labor, which continued until the enactment of the Labor Management Relations Act of 1947. The agency grew in size and scope of activity during the period and, with a relatively small field staff, achieved considerable success in settling industrial disputes. From 1913 to 1931, the service acted as mediator in some 10,000 disputes, with successes in over 70 per cent of the cases.

In 1947 this service was replaced by the Federal Mediation and Conciliation Service organized under a director responsible directly to the President rather than to the Department of Labor. The belief of Congress in collective bargaining as the most satisfactory method of settling industrial disputes is reaffirmed by the Taft-Hartley act, and thus every effort is made to make conciliation, mediation, and voluntary arbitration supplement rather than replace such bargaining. The conciliation service may be made available only in disputes that threaten an interruption in interstate commerce, and only as a last resort and in exceptional cases for disputes involving the application or interpretation of existing contracts. It is directed to furnish data and factual information and to use its best efforts to bring the parties in a dispute to an agreement by conciliation and mediation.

A special agency, the National Mediation Board, was created by the Railway Labor Act of 1926 to mediate industrial disputes on the railroads. If the board is unable to settle a dispute, and if the parties refuse to submit the dispute to arbitration, the board may

report to the President that an emergency exists. The President may then appoint an emergency fact-finding board. During these proceedings, and for thirty days after the report of the emergency board, neither party to the dispute may take aggressive action. Between 1926 and 1946 there were no important railroad strikes, partly because of the existence of the board, but also because of personal intervention by the President in almost every major labor dispute.

This twenty-year record was broken in 1946 in connection with a threatened strike by two of the railroad brotherhoods. A firm attitude by the President of the United States, strongly supported by public opinion, finally forestalled the strike, after mediation, arbitration, and "fact-finding" had failed.

The National Labor Relations Board, while primarily a law-enforcement agency, also does much essentially mediatory work. During preliminary investigation of unfair labor practice charges, it is frequently possible to induce a union to withdraw unfounded charges or to induce an employer who has been violating the act to comply voluntarily with its terms. About 95 per cent of the complaints filed with the Board are disposed of in this way, and formal proceedings leading to a cease and desist order are undertaken in relatively few cases.

Most of the states have added mediation to the list of duties of their industrial commissions or departments of labor.

Voluntary arbitration: General nature. Arbitration involves the submission of a dispute to an impartial umpire, who will survey the merits of the case and render a verdict. Voluntary arbitration usually presupposes an existing agreement which makes provision for the employment of an arbitrator, provided both parties agree. The award is binding on neither party, since it is nonenforcible at law.

However, when an agreement contains an arbitration clause, it is extremely difficult for either party to refuse later to submit a cause in dispute without loss of public support, and continued practice of arbitration gradually builds up a moral sanction which has great force in controlling conduct. The very fact that all previous steps have also been voluntary and that the arbitrator has been chosen by both sides because of his impartial character gives the award a sanction which is often more binding than a judgment based on force. Experience has shown that few unions or employers repudi-

ate such arbitration awards, provided the arbitrator has not overstepped the bounds of his mandate.

Official arbitration in the United States. Almost all governments have set aside public officials or boards to act as arbitrators on request of the parties to a dispute, either as an incidental feature of their duties or as their sole function. Officials or boards empowered to offer their services as mediators may also in most cases serve as arbitrators. Virtually every one of the American states has made some such provision for voluntary arbitration. The Erdman and Newlands acts provided that the designated mediation officials were to act as arbitrators of disputes on the railways upon consent of both parties if conciliation failed to compose the controversy, and the other federal laws providing for mediation have made their machinery available for voluntary arbitration.

On the whole these ex officio arbitrators have had little success either in this country or elsewhere. Both sides are inclined to distrust the impartiality and the capacity of men who owe their positions to political fortune. An improvement on the system is the appointment of a panel from which arbitrators may be chosen by the two parties, rather than the designation of a specific official. This is done in some twenty American states, whose laws provide for the appointment by the governor of representatives on the Board of Arbitration from nominees of employers and workers.

Effectiveness of voluntary arbitration. Aside from questions of personnel, the method of voluntary arbitration is of limited effectiveness, both as regards the range of the industrial area in which it can be made to operate and the type of dispute to which it can be applied with hope of success. It is practicable only in highly organized industries where trade unionism and collective bargaining are well established.

Even here it works only if the arbitrator has to deal with definite, concrete problems, such as an increase or decrease of wages or an alteration of hours. Controversies which raise really fundamental questions, such as the recognition of the union or the employer's right of discharge, are considered so vital to the interests of one side or the other that they will not be submitted to arbitration. Moreover it is not to be expected that an organized group will favor this method of settlement if the net result is to deprive it of the advantages of a strong bargaining position.

In his handling of matter-of-fact problems, therefore, the arbi-

trator must generally discard principles of abstract justice and right and be content to adjust his award to the relative strength of the contestants, attempting to give the stronger party the gains it would have made through an appeal to force. At best then arbitral awards are not deductions from economic law or from premises of public interest but a method of giving a potential victor the spoils of war without the expense of battle.

Transition stages. There are several transition stages between such voluntary arrangements as have been described and a fully developed system of coercive public intervention. One intermediate stage is represented by laws which make submission of disputes to arbitration compulsory, while leaving the acceptance of the award voluntary. Another stage is reached when attempts are made to give the award binding force after arbitration has occurred through mutual consent of the two parties. In an informal way this is sometimes done by requiring the parties to an arbitration agreement to deposit sums of money, which are forfeited in event of refusal to accept the decision of the arbitrator. The laws of about a dozen American states provide for punishment, by fine or imprisonment, of a recalcitrant party who has previously agreed to arbitration by public officials. These laws have proved impossible to enforce, and their presence on the statute books has acted as a deterrent to voluntary arbitration.

Compulsory investigation. A variant of this quasi-compulsory system is the prohibition of strikes and lockouts for a stated period while the authorities of the state investigate the matter and publish their findings. The best example of this system of *compulsory investigation* is to be found in Canada, where it has been in operation since 1906.

The Canadian law applies only to public utilities, transportation agencies, and mines. It provides that any change in wages and hours must be preceded by thirty days' notice. If the change is unacceptable to one of the parties, no strike or lockout may occur until after investigation of the facts by a board appointed by the Minister of Labor and public announcement of a decision by the board. Fines may be imposed for illegal strikes or lockouts during the investigation period. After the announcement of the award, however, both sides are at liberty to resort to a trial of strength. Experience has shown that, though the Canadian act has not stopped strikes entirely, it has been successful in 90 per cent of the cases to which it

has been applied. This record of success has been overshadowed by the fact that certain of the illegal strikes have been of large scale, causing much loss to the public.

The system has brought out one significant fact; *i.e.*, that the coercive features of the law are much less effective than its voluntary features. In fact the attempts to impose legal penalities on recalcitrant workers have proved so futile that this provision of the law has become inoperative. During the first ten-year period only eleven persons were prosecuted, although upwards of 80,000 had broken the law, and only $1,660 were collected in fines as compared with $150,000,000 as the maximum legally due.

Although twenty-five of the American states have conferred powers of compulsory investigation upon their departments of labor, only Colorado has given this procedure extensive trial. In this state a system very similar to the Canadian law was set up by the Industrial Commissions Act of 1915. This act was passed at the end of a disastrous strike in the mines.

The scope of the law has been altered at various times, but since 1923 it has been limited to the public utilities, intrastate railways, and mines. It copies from the Canadian law the requirement of thirty days' notice of change in working conditions, the prohibition of strikes pending investigation, and the public announcement of a nonmandatory award by the Industrial Commission. The period since 1915 has been on the whole one of industrial peace in Colorado, but it is impossible to say to what extent this was due to the operation of the law or to a system of conciliation set up at the same time by employers and workers.

A form of compulsory investigation is relied upon by the Taft-Hartley act to bring about the settlement of any threatened or actual strike or lockout affecting an entire industry, or any substantial part thereof, engaged in interstate or foreign commerce and imperiling national health and safety. In such cases, the President may appoint a board of inquiry to investigate and make a preliminary report on the dispute. When this report is received, the President may arrange to have filed a petition for the issuance of a temporary injunction effective for sixty days, during which the board of inquiry is required to reconvene to investigate and report upon the current position of the dispute and the efforts that have been made to settle it. Within fifteen days after the submission of the board's second report, the NLRB must take a secret ballot of

the employees involved on the acceptance or rejection of the employer's last offer and report the results thereof to the Attorney General. Thereupon the Attorney General is required to arrange for the discharge of the injunction, and the President is required to submit a full report of the dispute to Congress.

On the whole, organized labor opposes compulsory investigation, chiefly because the enforced delay weakens their power, in case a strike eventually occurs, by giving the employers time to make preparations for the strike.

Compulsory arbitration: The affirmative argument. In its complete form, compulsory arbitration prohibits strikes and lockouts entirely and enforces the arbitral award by governmental power. Many are convinced that the progress of law must lead eventually to this coercive intervention of the state in the case of all trade disputes. It is argued that the next step in the protection of the public from disorder must be the complete disarming of capital and labor for private warfare. Founded on considerations of this character, compulsory arbitration was adopted by the Australasian republics, and for a time by one American state, Kansas. The dictatorship countries uniformly prohibit strikes and require settlement of disputes by official decree. These procedures however have no pertinence to a social order based on freedom and democracy.

Experience in New Zealand. New Zealand was the first country to adopt a system of compulsory arbitration. The first law was passed in 1894, after a period of bitter strikes. Since that date, except for the years 1932–36, there has been an unbroken history of compulsion in New Zealand.

The original system had as its basis a series of regional conciliation commissions covering the entire island, each commission under the direction of a salaried expert. Attempt was made to settle all grievances by voluntary adjustment with the aid of these commissions. Failing this, an appeal from one of the parties took the case before a Court of Arbitration, which had final jurisdiction over all disputes.

Until 1914 strikes and lockouts were not illegal unless application had been made for an award by the court, but on and after such application all strikes by unions registered under the act were prohibited. Registration by the unions was voluntary, and those which did not register were not denied the right to strike. However the ease with which an entire labor group could be registered by a

minority of its members gave the system of compulsion wide sway. A law of 1913 set up a scheme of compulsory investigation, similar to that of Canada, applicable to unregistered workers. This law prohibited strikes during the investigation period.

The Court of Arbitration consisted of a president, taken from among the justices of the Supreme Court, and two other members chosen from nominees of the registered unions and the registered employers' associations respectively. The court had judicial power to compel the attendance of witnesses and the submission of books and documents. Violation of awards was punishable by fine, and each award held good for three years unless superseded by a new judgment of the court, even though the union to which it applied had canceled its registration in the meantime. Awards issued in cases affecting a single enterprise could be made applicable to the entire industry, thus equalizing competitive conditions.

The employers in New Zealand finally united against the law. Successful in a political campaign in 1932, they amended the arbitration procedure to remove most of its compulsory features, providing that, except for unions composed chiefly of women, submission to arbitration should require consent of both sides of the controversy. In other words, New Zealand returned at that time to a system of conciliation with voluntary arbitration as a last resort.

The change was of short duration, for the election of 1936 brought a labor government into power, which reinstated the compulsory features of the law and widened its scope. Arbitration is once more obligatory, and awards may be extended to cover all firms in a competitive area. The new legislation also makes union membership compulsory for workers eighteen years of age or over who are receiving the adult minimum wage rate. A union which strikes in violation of the arbitration system, however, may be punished not only by fines but by cancellation of the requirement for compulsory union membership.

Compulsory arbitration in Kansas. The Kansas law of 1920 is the only example, outside of Australasia, of thorough-going compulsion exerted by a democratic government. This law prohibited strikes and lockouts in specified industries and created a Court of Industrial Relations to settle all disputes.

Early in 1925, however, the court was shorn of most of its power by a decision of the Supreme Court of the United States, which declared unconstitutional the compulsory arbitration features of the

law as applied to working time and by implication overthrew the whole process of compulsory arbitration whenever its effect was "to compel owner and employees to continue the business on terms not of their own making." This decision indicates the apparent unconstitutionality of all measures seeking to apply the principle of compulsion to privately owned enterprises in this country, except in cases of emergency in industries clearly affected with public interest, as long as the federal constitution contains a clause guaranteeing "liberty of contract."

The United States War Labor Board. The first important American venture in compulsory arbitration on a national scale was undertaken during the Second World War. At a conference called by the President in December, 1941, leaders of labor and management pledged themselves to refrain from strikes and lockouts during the war and to submit unsettled disputes to a national war labor board. This board, created by the President early in January, 1942, and later legalized, consisted of four labor, four management, and four public members.

Through its own decisions and those of its regional boards and industry commissions, which were also tripartite bodies, the board during its four years of life dealt with many thousands of disputes. Board decisions were binding on the parties. In the few cases of refusals to comply with a board award, the award was enforced by seizure of the plant and its operation by a government agency. While strikes in war industries were not eliminated, they were reduced to a very low level, and there can be no doubt that the board was an essential part of the wartime economic administration.

The most frequent criticism of the board from labor involved the slowness of its procedures. Cases frequently dragged on for a year, and sometimes for as long as two years, before final settlement. The tripartite composition of the board, while useful in securing acceptance of its decisions, did not enable labor and management groups to make the decisions. The policy of the board was the policy of the public members, who had the deciding vote on all disputed issues. It also became increasingly apparent that compulsory arbitration involved government determination not only of wage rates but of all other important matters of union-management relations.

Under the pressure of war needs labor and management reluctantly accepted this situation, but their reluctance grew as the war

progressed. As soon as the war ended they withdrew their support entirely, and the board was abolished in December, 1945.

Appraisal of compulsory arbitration. Any inferences regarding the practical operation of compulsory arbitration must be drawn from the experience of Australasia, where this method of promoting industrial peace has had thorough trial. For a time the world was inclined to accept the evidence as proof of success. New Zealand was known until 1905 as a "land without strikes." But the subsequent history of the system has shown that strikes cannot be prevented by compulsion. Since 1906 New Zealand has experienced frequent and bitterly contested struggles in the field of industrial relations. Australia has had similar experience.

The early success of these laws was due chiefly to the contentment of the workers with a system which operated continuously in their favor. The Courts of Arbitration favored the principle of unionism, even to the extent of compelling employers to accept the closed shop. Early awards resulted in continuous advance of wages and shortening of hours, thus conceding to labor larger material gains than could have been obtained from the process of bargaining alone. There were no strikes, because the causes of strikes were absent.

The underlying reason for this result favorable to labor was the rapid and sustained material progress of Australasia during this early period of its industrial development, which made possible a continued increase of the wage income without encroaching on the necessary rewards of the other factors of production. When the rate of progress was retarded and it became impossible to grant the demands of the workers, rebellion against the system arose immediately. An epidemic of strikes broke out, which were the more bitter because they were levied in part against a government which was thought to be unfair to labor.

When the workers used their political power to influence the process of arbitration, the result was a series of decisions which laid unsupportable costs on the business world. In New Zealand owners of industry were then driven in their turn to take political action against the system, with the result, as we have seen, that it was deprived of its mandatory features. The political victory of labor which restored compulsory arbitration in 1936 added new emphasis to the fact that the state exercising such powers is destined

to become the centre of a political struggle between capital and labor.

No type of public intervention has as yet proved as effective in promoting industrial peace as have voluntary arrangements set up within industry by mutual consent of employer and worker. Conciliation, mediation, and voluntary arbitration are accomplishing much in the way of smoothing out labor disputes. Machinery for public intervention in the last resort is most useful when it is carefully designed to promote these informal devices for negotiation and settlement. But these voluntary arrangements, as we have seen, presuppose organization on both sides, and in the United States at the present moment there is no agreement within industry as to the mechanism through which organization can be brought about.

Attempts to apply compulsion instantly arouse the suspicion and antagonism of labor. If they are persisted in, despite this antagonism, their ultimate effect is to force organized labor into politics, with a program of far-reaching social change which might mean the end of the present system of industry.

CHAPTER XXXV

Problems of American Agriculture

Economic peculiarities of agriculture. Stated in its simplest terms, the farm problem arises out of the failure of competitive price to bring about easily and quickly the adjustments in production that would preserve a normal balance between costs and prices. In agriculture normal equilibrium is reached with more difficulty than in industry, generally because of the vagaries of nature's bounty, the inelasticity of demand which frequently prevails, the highly individualized control of productive efforts, the less compelling influence of marginal costs on productive capacity, and the greater inflexibility in the transfer of productive agents from use to use. The biologic processes involved in crop and livestock production require considerable periods of time; from three to five months elapse between the planting and harvesting of most crops, while several years are usually required to make important changes in crop rotations and livestock breeding programs. Even without the uncertainties that always accompany forecasts of future conditions, most farmers have very imperfect knowledge of present conditions and of the most economical methods of farm organization. Finally the typical farmer is far from being the "economic man," and many of his decisions are motivated or modified by conditions other than economic rationality.

Consider the effects of weather conditions on crop production. From his experience, the Maine potato farmer may expect that his current crop will yield from 230 to 290 bushels per acre, but he also realizes that since 1920 average potato yields in his state have ranged from 177 to 355 bushels per acre. His conscious efforts to direct production by varying the acreage planted is as often as not offset by conditions beyond his control.

The same holds true for other crops and other areas. Average corn yields for the state of Iowa in recent years have ranged from 20 to 60 bushels per acre while wheat yields in North Dakota have varied from 5 to 20 bushels

Yet the major uncertainty is not so much in production as in price. The year-to-year fluctuations in production are accompanied by more or less opposite and often more pronounced fluctuations in price. In the two decades after 1920, yearly average prices for the whole country have varied from 32 cents to $1.70 per bushel for potatoes, from 38 cents to $1.85 per bushel for wheat, and from 32 cents to $1.12 per bushel for corn. In addition there were important seasonal and geographical price differences. Such fluctuations make it difficult to distinguish between fundamental or long-run changes and short-run changes, and it is quite impossible to forecast with consistent accuracy the prices that will hold in the future.

The result is that most farmers make their production plans on the basis of current prices. A recent study of hog producers in Iowa indicated that 60 per cent forecast hog prices nine months in the future at the then current level of approximately $5.00 per hundred pounds.[1] The remaining 40 per cent made forecasts ranging from $3.50 to $7.00. Actually prices increased to approximately $6.00 and continued to increase rapidly as a result of war conditions.

A further complication is that many costs do not involve a direct cash expenditure. In an investigation of the costs of wheat production on an Iowa farm, it was found that about half of the costs were of the noncash type, including especially the labor of the farmer and his family and the return on his own investment.[2] Of the remaining 50 per cent, about 20 per cent represented costs such as home-grown feed and seed, the use of which, instead of sale at current prices, involved a direct sacrifice of cash income. Only 30 per cent of the estimated total costs of production were direct cash costs, and a substantial part of these were for such fixed items as taxes, interest, and payments on borrowed capital. In short, variable cash costs are a minor part of the total.

When prices fall and the income from a particular crop is insufficient to yield the usual market return on his labor and capital, the farmer who cannot shift profitably to the production of another crop will continue to produce the same crop so long as

[1] Schultz, T. W., and Brownlee, O. H., "Two Trials to Determine Expectation Models Applicable to Agriculture," *Quarterly Journal of Economics,* Vol. LVI, Number 3, May, 1942, pp. 487-496.

[2] Pond, G. A., *Cost of Production and Price,* University of Minnesota, Extension Bulletin 166, Revised, 1941.

the price is sufficient to pay anything more than the direct cash costs, even though he receives no return at all for the use of his fixed capital and a reduced rate of return for his own services. Rarely can he afford—immediately at least—to discontinue production and thus deprive himself and his family of a job, for there are few alternative uses to which farm land or farm buildings can readily be put. A long continued period of production at prices which do not cover costs, including the maintenance of the productivity of the soil as well as a fair return to the farmer for his own labor, will inevitably throw land out of production, but a decision to abandon production is in most instances reached most reluctantly and only after some years of adversity.

Even this brief discussion should make it evident that in agriculture responses to price changes are slow and imperfect. It is true that higher prices lead to higher production and, with more difficulty, that lower prices bring reductions in the output of particular products. But the uncertainties as to costs and future prices create important lags and imperfections in the process. Viewed solely as an economic problem, we may describe this as a failure of the market to bring about quickly and easily the optimum allocation of resources. But to the farmer involved it may mean long years of hard work and poverty, ending in loss of the savings of a lifetime.

The American farm problem: Period of the First World War. In the years immediately preceding the First World War agriculture was prosperous in a quiet sort of way. It had achieved a measure of stability, and except for the effect of weather on crops, production, prices, and farmers' incomes remained steady and in general satisfactory.

When war broke out in midsummer, 1914, agricultural prices were momentarily depressed, but by September foreign demand for certain farm products strengthened and exports began to increase. This demand, however, was far from uniform, for while wheat and hogs were in great demand, poultry products went begging. In general, exports were limited by the difficulties of storing and transporting foods, and only those products that could travel long distances and maintain their quality were wanted. Except for a few products like wheat, farm prices showed no increase over prewar levels until early in 1916.

Wheat prices, however, advanced sharply in the first year of the war, and under the influence of rising prices and good export mar-

kets, acreage expanded. The increase in acreage combined with unusually favorable weather in 1915 to yield a record wheat crop that was not surpassed until 1944. This big crop came at the same time that other wheat-producing countries had good crops, and it forced prices back to prewar levels.

After the United States entered the war in 1917, minimum prices were guaranteed for wheat, and every effort was made to increase production. These measures were successful, and wheat acreage again soared, reaching a peak in 1919. Wheat prices remained high until the harvest in 1920, when there began a swift decline that lasted through 1921.

The other farm product most conspicuously affected by the war was pork. Until 1916 hog prices did not change much. Then they started upward and did not stop until the summer of 1919. At that time, earlier than for most other farm products, the price broke, and a decline set in that lasted until prewar levels were reached in 1921. As with wheat, high prices and other incentives had their effects upon hog production, and the pig crop increased from 53 million hogs in 1914 to 64 million in 1919.

Wartime demands stimulated the production of many other agricultural products. All told, the area of land in crops was expanded by about 40 million acres, or 13 per cent.

An important development that accompanied rising agricultural prices and expanded production was the increase both in farm costs and in land values. Prices of fertilizer, machinery, farm labor, and the like rose, taxes on farm real estate and interest on farm mortgages more than doubled, while freight rates showed substantial increases. When everything was taken into consideration the increases in farm expenses were sufficient to offset in large measure the gains from increased prices of farm products. Nevertheless the increase in money income, together with the assumption on the part of many that high war-time prices would continue, whipped up a spirited speculative bidding for land, with the result that the average value per acre of farm real estate rose 70 per cent between 1913 and the beginning of 1920.

Not only did the farmers' money income rise during this period, but their real income as well. At the beginning of the war the prices of agricultural products rose more rapidly than did other prices, and there was a time of real prosperity for farmers, when the ratio of the prices they received for their products to the prices they

paid for goods was considerably more favorable than in the years 1909–14 (Figure 39).

Before the end of the war this situation began to change, for while farm prices still rose, other prices rose faster, and finally, in 1920, farm prices turned rapidly downward. The favorable ratio of prices received to prices paid was thus changed, and before the end of 1920 it had fallen far below the prewar level. This ratio was never regained until the Second World War.

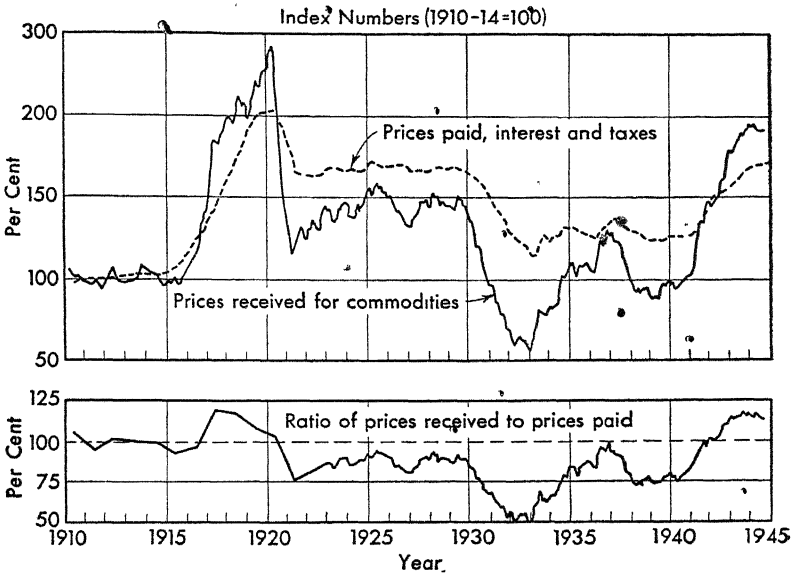


FIG. 39. PRICES RECEIVED AND PAID BY FARMERS IN THE UNITED STATES, 1910-45 [1]

During the 1920's. Thus it came about that farmers found themselves in the middle of 1920, when prices were at their peak, with no increase in real income, and with an expanded plant, inflated land values, and increased production costs. Since other prices showed greater rigidity, the precipitate break in agricultural prices late in 1920 caused a great disparity between farm and industrial prices. This persisted throughout the decade, despite a slight recovery in farm prices in the latter half of the period (Figure 39). Farming

[1] Figures obtained from Bureau of Agricultural Economics and Agricultural Marketing Service, *Agricultural Statistics*. Washington: Government Printing Office, annual report.

was unprofitable, and thousands of farmers suffered loss and hardship. Hardest hit were those with cash payments to make, and farmers who had bought land on credit at high prices found it difficult or impossible to pay interest and make repayments on farm mortgages. A farm problem of formidable dimensions existed and could not be ignored.

The failure of agricultural production to adjust quickly and easily to marked decreases in demand was strikingly evident in the period following the First World War. Throughout the 1920's the production of wheat, which may be taken as a sample, remained about 35 per cent above the average production of 1909–13.

In part the failure to reduce production may be attributed to the difficulties inherent in the system which have been described above, but the fact that leaders of agriculture seem consistently to have misread the situation was a contributing factor. In 1920, for example, editors of farm papers and other leaders appeared to have no misgivings regarding the export market for farm products, and throughout the following decade the feeling among both the leaders in farm affairs and the professional speculators was that the export market was bound to recover.

One authority writes: "The prevailing opinion, in both trade and government circles, was that wheat prices were abnormally low in 1928–1929; but in retrospect it is clear that speculative forces held them above levels appropriate to the real conditions. The danger signals given by unprecedentedly high 'visible supplies' of wheat were largely ignored . . . the conviction prevailed that 1929–30 would see higher prices and ready absorption of surplus wheat stocks. This was the situation when the Federal Farm Board came into existence in July, 1929, and when the financial crisis broke in the New York stock market in October [of that year]. In his annual report of 1929, written shortly after this, the Secretary of Agriculture observed that the agricultural situation was remarkably free from evidence of overproduction." [1]

A second reason which may be given to explain the continued maladjustment in agriculture is the conviction throughout the period that prices were unjustly low and the expectation that Congress would do something for agriculture. Constant pressure by farm groups was exerted, and a number of measures supposed to

[1] Davis, J. S., *Wheat and the AAA*. Washington: The Brookings Institution, 1935, p. 17.

aid agricul ure were either enacted into law or passed by Congress, only to be vetoed by the President. It is not our purpose here either to describe these different projects or to pass judgment on the plea of agriculture for special treatment at the hands of the government, but merely to record the fact that a decade of political agitation and political promises could not fail to influence the policy and behavior of the farmers.

During the 1930's. Farm prices and income Even without the depressing effects of the world crisis, American agriculture would have had difficulty in solving the problems that confronted it in the years following 1929. The world crisis, however, put an end to American loans to foreign countries, thus causing collapse of the export market for American produce, while internal disturbances wrought havoc with the domestic market. Prices declined generally, but farm prices fell even more rapidly, so that there were drastic reductions, not only in the size of the farmer's income, but in its purchasing power.

GROSS FARM INCOME AND NATIONAL INCOME, 1910–1945[1]

Year	Gross farm income[a] (Billion dollars)	National income (Billion dollars)	Income parity ratio[b]
1910	7.4	33.1	105
1919	17.7	66.1	152
1921	10.5	58.3	60
1929	13.8	86.0	93
1932	6.4	43.6	61
1937	11.3	71.6	107
1940	11.0	78.5	90
1941	13.8	94.8	107
1942	18.4	122.2	134
1943	22.8	149.4	150
1944	23.9	160.7	145
1945	24.6	161.0	

[a] Including government payments.

[b] Obtained by dividing the index number of income from farming per person on farms by the index number of income per person not on farms. This provides a measure of the relation of the net income from farm operations of individuals on farms to the income of individuals not on farms, as specified in the Agricultural Adjustment Act of 1938.

[1] Sources: Bureau of Agricultural Economics, *Net Farm Income and Parity Report*, 1945, and *The Farm Income Situation* (Series). Also, Bureau of Foreign and Domestic Commerce, *Survey of Current Business* (Series). All the foregoing published in Washington by the Government Printing Office.

The fall in the prices of farm products was naturally reflected in a shrinkage in farm income, as shown in the table on the preceding page. Gross farm income, estimated at 13.8 billion dollars in 1929, fell to 6.4 billion in 1932, a decline of 54 per cent in three years. By comparison, the total income of American enterprise of 86.0 billion dollars in 1929 underwent a reduction of 49 per cent by 1932.

Farm debt. The farmer's situation was aggravated by an unusually heavy burden of debt, estimated at a total of 12 billion dollars in 1933. Part of this consisted of short-term loans, but the major portion was in the form of farm mortgages, which were to a considerable extent the result of purchases of farm lands at the speculative prices obtaining during and immediately after the First World War. This mortgage debt had reached a peak of 10.8 billion dollars in 1923 and was not reduced very much until after 1930.

The interest on this debt amounted in 1929 to about half a billion dollars yearly. With farm income curtailed, many farmers found themselves unable to meet their interest payments, and foreclosures followed. Between 1920 and 1930 nearly half a million farmers lost their farms through foreclosures and forced sales, and many formerly independent farmers were forced into the class of tenants. In 1910 nearly two-thirds of all farmers were owners, whereas in 1932 over half were tenant farmers. Between 1930 and 1934 the farm mortgage debt decreased from 9.6 billion dollars to 7.7 billion dollars through continuation of the process of foreclosure and sale, but even so not all of the delinquent mortgages were foreclosed. Mortgage moratoria, some of them later invalidated by the courts, were established by state law in many of the agricultural states, and in their absence organized mass movements by farmers interfered with court proceedings and halted foreclosures and forced sales.

Federal Farm Board. During the decade of the 1920's, the farmers made insistent demands for legislative relief for their economic difficulties. Little in the way of effective help was obtained, however, until the passage of the Agricultural Marketing Act of 1929. This act set up a Federal Farm Board, authorized to encourage and strengthen coöperatives and to engage in stabilizing operations.

The board had been in existence but a short time when the market crash of 1929 made extensive stabilization operations seem imperative. Beginning in 1930 the grain and cotton stabilization corporations of the Federal Farm Board did considerable buying of wheat and cotton.

From July, 1930, to August, 1932, over a million bales of cotton so bought were being held. The bulk of this was distributed, after processing, through the National Red Cross, and the remainder, together with one and one-half million bales which were held as a result of collateral loans, was turned over to the Department of Agriculture in 1933 for the cotton option pool under the new farm legislation.

As to wheat, the accumulated holdings of the Farm Board were liquidated in 1932 and 1933, largely by distribution through the National Red Cross after conversion into flour. By the end of April, 1933, the Grain Stabilization Corporation had closed out all its holdings, both spot and future.

The Federal Farm Board came to an end in May, 1933, with its usable credit functions salvaged for the new Farm Credit Administration. Heavy surpluses of wheat and cotton hanging over the market and the failure to offer any discouragement to increases of production had helped to create an almost continual decline in the prices of these commodities during the life of the board. Nor did the board's operations tend to decrease wheat production. The final loss on the 500 million dollar revolving fund for price stabilization was about 350 million dollars.

Farm programs of the Roosevelt administration: General objectives. The purpose of the government's farm program in 1933 can be simply stated. It was to bring, with as little delay as possible, the prices of farm products into a more favorable balance with the prices of other goods and to stabilize this more favorable relationship once it had been attained. This was essentially the objective set forth by the McNary-Haugen and other abortive farm relief plans of the past.

The agricultural crisis was deemed so serious that various means for achieving this objective were tried simultaneously. In so far as the program for raising the general price level was based on the assumption that, once prices started moving upward, the more depressed prices would make the greatest relative recovery, the general price program should be considered a part of the farm program. New agricultural credit agencies, liberally supplied with government funds and under government control, were established to make loans to the farmers and thus to give them sufficient financial strength to hold their produce for a favorable market. Devices for enlarging foreign markets, such as the negotiation of reciprocal

tariff agreements authorized by the tariff act of 1934, or the export of agricultural produce with government subsidy, were also directed toward this same end.

The Agricultural Adjustment Act of 1933: Objectives. The main program, however, sought (1) to make it immediately profitable for producers of the basic agricultural commodities to reduce production voluntarily, (2) to check the development of unfavorable competitive conditions, and (3) to supervise the growth of equitable distributive processes for all agricultural commodities. This program was embodied in the Agricultural Adjustment Act of 1933 and its various amendments. The 1933 act was in effect for less than three years, being declared unconstitutional in January, 1936, but in that period the administration was able to exercise considerable control over American agriculture.

Enduring changes in the pattern of American agrictulture and in the point of view of the American farmer could not fail to result. For years a steadily increasing number of farmers and farm leaders had agitated for a governmental program of this kind. The economic philosophy that gave rise to the drafting of the act of 1933 can be traced through the McNary-Haugen bills of the 1920's and far back into the agricultural programs of the nineteenth century. Furthermore it did not die with the court decision in 1936. It persisted and found tangible expression in subsequent farm legislation. A critical study of the act of 1933 is, therefore, of more than historical interest. It provides a most essential background for an understanding of the objectives of the federal government and for any evaluation of the probable results of the present laws affecting agricultural production.

"Parity prices." The Agricultural Adjustment Act was entitled "An act to relieve the existing national economic emergency by increasing agricultural purchasing power," etc. Its major goal was restoration of the purchasing power of basic farm products to the level of the prewar period—August, 1909, to July, 1914—except in the case of tobacco and potatoes, for which the base period was from August, 1919, to July, 1929. Originally the list of basic farm products consisted of wheat, cotton, hogs, field corn, rice, tobacco, and milk and its products, but the number was extended by later legislation.

For each basic commodity there was at any given time a so-called "parity price." For example in the base period the average price

of wheat was $0.884 a bushel. In order that a bushel of wheat should have as great a command over commodities in general as it had in the base period, the price in April, 1936, for example, would have to be $1.105. In this same month the parity price for cotton was 15.5 cents a pound, compared with 12.4 cents in the base period, and so on for the other basic commodities. The plan did not attempt to set fixed prices but sought to adjust farm prices to the general price level, whatever that might be.

Production control. In order to limit production, the Secretary of Agriculture was empowered to provide, by voluntary agreements with producers, for the reduction of acreage or production for market of the basic commodities and to compensate the coöperating producers by the payment of rent, if the land was leased out of production, or by benefit payments on the restricted allotments grown. Variations of this general scheme were used for cotton. In later measures voluntary crop reduction for cotton was replaced with compulsory reduction, and compulsory control was also made applicable to sugar, tobacco, and potatoes.

In addition the Secretary of Agriculture was permitted to make marketing agreements, exempt from the operation of the antitrust laws, with processors and handlers of any agricultural commodity or its derived products in the current of interstate or foreign commerce."

By the original act of 1933 the Secretary of Agriculture was empowered to use the proceeds of all taxes authorized by the act for the purpose of expanding the foreign market for American produce and thereby removing the surplus from the domestic market. It was by virtue of this authority that an agreement was made in 1933 for the export of surplus wheat from the Pacific northwest by the North Pacific Emergency Export Association. An average loss of 22 cents a bushel was made good out of funds derived from the processing tax described below.

An extension of this power to develop foreign markets is found in the amendment of 1935, which permitted the use of 30 per cent of the receipts from import duties to subsidize the export of farm products, with the exception of raw cotton, or to increase the returns of producers on the part of the crop consumed in this country.

Processing taxes. The cost of the payments mentioned above was to be met from taxes on the first domestic processing of the products. The Secretary of Agriculture was given the power to set

the amount of the tax on each basic commodity at any figure up to the amount of the difference between the current average farm price of the commodity so processed and the parity price. The processing taxes were remitted on products exported and were not levied on supplies processed by the grower for his own consumption or used for charity. A compensating tax was levied on any imported manufactured article whose value was chiefly derived from one of the commodities concerned.

Wheat program. Under the wheat program farmers were offered money "benefits" to induce them to agree to stipulated reductions in the area to be sown to wheat. In this way it was hoped the production of wheat would be so reduced that the effective return per bushel of wheat would have, for the contracting farmers, a purchasing power over other commodities more nearly in accord with that obtaining in the base period.

So far as the reduction of acres sown to wheat was concerned the program was not successful. The 1934 program called for a 15 per cent decrease in acres planted to wheat by farmers signing contracts with the government. Despite the fact that farmers controlling 75 per cent of the wheat acreage of the base years, 1928–32, signed these contracts, increased plantings by nonsigners resulted in a net decrease in 1934 of 4 per cent only. The 1935 program called for a reduction in acreage of 10 per cent by signers, whereas actually the acreage planted was in excess of that for the base years. Production did decline, however, from an average of 864,000,000 bushels in the base period to an estimated yield of 496,000,000 bushels in 1934 and 626,000,000 bushels in 1935—crops small enough to cause a rapid decrease in the large carry-over of wheat in 1933.

This might be considered evidence of success of the program if one failed to remember that unfavorable weather conditions prevailed during these years. The investigation of the Brookings Institution indicates that had there been no attempt to control production the 1934 crop might have been 25 to 30 million bushels, and the 1935 crop perhaps 20 million bushels larger, than they actually were.[1] In other words the effects of the Agricultural Adjustment Administration measures of restriction were practically negligible in comparions with the effects of weather conditions.

[1] Nourse, E. G., Davis, J. S., and Black, J. D., *Three Years of Agricultural Adjustment Administration.* Washington: The Brookings Institution, 1937, p. 127.

Other programs. The wheat program was, as has been pointed out, only one of the programs undertaken under the AAA. Controls were applied to cotton, corn, hogs, sugar, and tobacco, and, through marketing agreements, to a number of other products. These programs varied in detail, but in general principle they followed the same pattern as that for wheat.

There can be no question of the accomplishment of the AAA in effecting a marked reduction both in the area planted to cotton and in the quantity of cotton produced. The program also added substantially to the income of cotton producers. But despite the success of the AAA in achieving its immediate goal, the program has been widely criticized. The most widespread objection has been that the restriction program and the policy of making generous loans "without recourse" to farmers on farm-stored cotton pledged as security have combined to reduce exports of United States cotton and permanently to endanger her export trade because of the maintenance of domestic prices above the world level.

AAA program and agricultural recovery. During the period of control, agricultural prices rose, and the ratio of prices paid to prices received by farmers moved closer towards equality. Gross farm income rose from 6.4 billion dollars in 1932 to 11.3 billion dollars in 1937, while the income parity ratio increased from 61 to 107. Carry-overs of some crops were greatly reduced. In general, agriculture was regaining her health.

For this improvement in the condition of agriculture the AAA cannot claim full credit. The farm value of the controlled crops was 85 per cent greater in 1934 than it had been in 1932, but the farm value of the noncontrolled crops was 50 per cent larger. A part of this remaining differential in favor of controlled crops is to be accounted for by the peculiarly severe effects of the drought. In the matter of area reductions also the AAA can take only partial credit for the changes effected, for while it is true that the acreage of controlled crops harvested declined 23 per cent between 1932 and 1933, individual adjustments to market conditions on private initiative were sufficient to produce a 13 per cent reduction in these years in the harvested area of noncontrolled crops.

AAA program and business recovery. The AAA program was designed not only to relieve the acute agricultural distress, but also as an integral part of the more comprehensive recovery program to benefit the whole nation. In fact the chief justification for the

AAA program lay in the argument that, if the farmers' buying powers were restored, orders would begin to flow to industrial centers and employment would pick up. This in turn would mean a greater demand for farm products and so on until more prosperous conditions generally had been brought about.

The influence of increased farm buying on general recovery is difficult to appraise. The general conclusion of the Brookings report to which we have referred above is that the AAA program had an appreciable effect on recovery. This resulted from an enlargement of total purchasing power (chiefly from government subsidies and loans) and not merely a transfer from one group to another, an acceleration in the rate of turnover of the funds involved, and a considerable advance in the time at which this purchasing power was made available.

Soil Conservation and Domestic Allotment Act. When the use made of the processing taxes was declared unconstitutional in 1936, it did not end the activities of the Agricultural Adjustment Administration. Not only was new legislation enacted shortly thereafter, permitting the Administration to seek the same objective in a different manner, but many of the powers conferred by the act of 1933 and its amendments were unaffected by the court decision. The Soil Conservation and Domestic Allotment Act was quickly passed, and this, together with the remnants of the adjustment act, provided the basis for the operations of the adjustment administration until the passage of the act of 1938.

Under the soil conservation legislation, the Secretary of Agriculture made payments to farmers who agreed to divert land from soil-depleting crops to soil-conserving crops, such as grasses and leguminous crops, and to follow soil-building practices. The program for 1938 and following years provided for a national goal in terms of acreage for soil-depleting crops and a national objective for soil-building crops and practices. The national goal was subdivided into state, county, and individual farm goals. There were individual soil-depleting crop goals for cotton, corn, tobacco, peanuts, rice, and potatoes. Other crops were classified together in a general soil-depleting crop goal. A maximum payment was calculated for each coöperating farm, and payments were made for keeping within soil-depleting crop goals and for attaining soil-building goals. Payments to farmers under this act came directly from the Treasury and were raised either by taxation or by borrowing.

Agricultural Adjustment Act of 1938: Reasons for enactment. The soil conservation act was never considered by the government or the farmers as a satisfactory substitute for the original farm legislation, since it gave relatively little opportunity to control the production of various crops. Its chief feature, aside from the promotion of soil conservation, in which the administration took a genuine interest, consisted in the improvement of farm income through the payment of subsidies from the public treasury.

In the circumstances it was inevitable that an attempt should be made to pass further legislation, which would not be deemed unconstitutional, giving the federal government comprehensive powers of control over agricultural production. This found expression in an act approved on February 16, 1938, the most important provisions of which are summarized in the following paragraphs.

Principal provisions of the Act of 1938: (a) *Soil conservation.* The act continued and supplemented the soil conservation program of 1936. The Secretary of Agriculture was authorized to prescribe bases for acreage allotments to carry out the program. Furthermore when Congress should provide the funds, payments were authorized to producers of corn, wheat, cotton, tobacco, and rice in such amounts as might be necessary to restore the 1909–14 ratio of farmer's incomes to the incomes of others.

(b) *Commodity loans.* The Commodity Credit Corporation was authorized to make loans on agricultural commodities. Such loans were made mandatory on wheat, cotton, or corn whenever the stock (as estimated on a specific date) should exceed a normal year's domestic consumption and exports. The corporation must also make loans on wheat if on July 15 of any year its farm price is below 52 per cent of the parity price, and on cotton if on August 1 the average price is below 75 per cent of parity. Loans made to coöperators in the soil conservation program should be at a rate between 52 and 75 per cent of parity, and loans to non-coöperators at a rate equal to 60 per cent of the rate to coöperators.

(c) *Parity payments.* If and when appropriations are made by Congress, payments may be made to producers of corn, wheat, cotton, tobacco, and rice to bring the effective return from the commodity up to its parity price.

(d) *Marketing quotas.* The Secretary of Agriculture is required to make known his findings with regard to stocks of commodities and to proclaim a marketing quota if the total supply of any of the

five commodities exceeds a given percentage of the supply for a normal year's domestic consumption and export. No quota is effective if more than one-third of those voting in a referendum oppose it. Acreage allotments are to be made for corn, wheat, cotton, rice, and tobacco on the basis of average production per acre over certain preceding years, and are to be so fixed as to produce an amount sufficient to supply domestic needs. Commodities marketed in excess of individual farm quotas are subject to definite penalties for each of the five commodities.

(e) *Crop insurance.* A government-owned Federal Crop Insurance Corporation was established within the Department of Agriculture, with capital stock to amount eventually to $100,000,000. This corporation is authorized, beginning with the 1939 crop, to insure wheat producers against loss in the yield of wheat from unavoidable causes, such as drought, flood, insects, plant diseases, and the like. Contracts must cover not less than 50 per cent nor more than 75 per cent of the recorded yield for a representative base period. Premiums are payable in wheat or the cash equivalent.

General objectives of the act of 1938: Financing. Perhaps the most conspicuous point of difference between the 1933 and 1938 acts was that the 1938 act omitted the processing taxes and in their place provided that benefit payments be paid when and if Congress provided the funds. The question as to the extent to which Congress wished to subsidize agriculture was thus presented annually. Congress could if it chose put an end to the entire program by the simple expedient of failing to vote the necessary funds.

Income parity vs. price parity. Income parity was set up as the ultimate objective in place of price parity. Price parity was still to be used as a measuring rod, but the goal now was to secure, through a rise in the prices of agricultural commodities and the payment of benefits, a net income for the farmer which would bear the same ratio to the income of the nonfarmer that it had in 1909–14.

The "Ever-normal Granary." A further objective of the 1938 act is best expressed by the term "ever-normal granary." This concept implied a steady flow of farm products to the market at prices which were reasonably stable over the years and sufficiently remunerative to the farmer to achieve the major goal of the program; namely, income parity.

The acreage allotments were supposed to be set at a figure which

would produce the desired quantities of the products, but these allotments were binding only on those farmers who chose to coöperate with the government. Therefore marketing allotments were provided, which in the event of embarrassing surpluses would make the marketing of produce in excess of the allotments subject to fines. Marketing allotments were compulsory acreage allotments, affecting all farmers when ratified by a two-thirds vote of the farmers concerned.

The administration recognized, however, that weather conditions, insect pests, human frailty in living up to quotas, and indeed errors in estimating accurately both the requirements for consumption and the acreage necessary to produce the amount might result in crops which were overabundant in one year and deficient in another. The commodity loan was specifically intended to even discrepancies between the requirements for consumption and export and the quantity produced in any given year. In time of glut and low prices the government was prepared, through the Commodity Credit Corporation, to make loans to farmers on produce stored on the farms or in commercial warehouses. The withdrawal of these stocks from the effective supply would therefore raise prices—presumably to or near the desired level. When a poor crop came along and prices would otherwise be higher than desired, the produce in storage would be released, thus adding to the effective supply and giving assurance of normal quantities for consumption and normal prices.

Crop insurance. Crop insurance was another innovation which was expected to work in the same direction, although it was intended also to relieve the farmer of some of the risk which had traditionally been his. Premiums were payable in wheat, the only crop for which insurance was originally provided, thus to some extent decreasing the quantity on the open market. If there were no occasion to pay out benefits, the wheat would remain in storage, but if a drought or a plague of insects destroyed a portion or all of a crop, the stored-up wheat would be released for the payment of insurance benefits, with the idea of giving the farmer a substantial portion of his normal income and assuring to the consumer supplies of wheat at prices close to normal.

Crop insurance began with wheat in 1939, but new legislation in 1941 provided insurance for the cotton crop also. Unfavorable financial experience with the program led Congress in 1943 to cut

off all funds for its support, save an amount necessary to liquidate outstanding commitments. Legislative authority for crop insurance was, however, not revoked, and in 1945 it was revived, with coverage of wheat, cotton, and flax, and with authority for experimental coverage of corn and tobacco.

Disposal of surplus commodities. Despite its past and potential importance, the government program for the removal of surplus farm products has been mentioned only briefly. The reader will recall the liquidation, early in the 1930's, of the surpluses accumulated under the price stabilization program of the Federal Farm Board. Since that time the program has consisted in part of subsidized export of surplus products. To some extent, however, funds have been expended to discover new uses and new domestic markets for farm products, and to provide free supplies of food for needy persons.

The idea of the latter was to lift the level of nutrition in the nation. This was an important consideration in the program for donating surplus food supplies for free lunches for school children and in the food stamp plan, although the improvement in the market for foodstuffs which might be expected to result therefrom remained the major objective. The food stamp plan was indeed based on the assumption that needy families could not themselves secure enough food to maintain a desirable standard of health and well-being; free blue stamps, giving buying power in retail stores over specified surplus foods, were accordingly furnished to families on relief.

During the war years of 1940–45, the number of people on relief declined and the growing need for food for the United States and its allies used up the surpluses in most lines. This phase of farm relief therefore receded into the background, though the program for school lunches was in 1946 placed on a permanent basis and given an enhanced appropriation.

Other agricultural programs. Production control and surplus disposal by no means exhaust the list of federal and state agricultural programs that were developed or expanded in this period. As a part of the policy of soil conservation the Soil Conservation Service was established to develop such programs as contour planting and terracing to control soil erosion, and a good deal of progress has been made. In 1943, sixty different soil- and range-building practices were being carried out on more than 3½ million farms.

Credit for farmers and for farm coöperatives was made available through the various activities of the Farm Credit Administration. This was supplemented by the Farm Security Administration with its programs of farm planning and liberal credit for low income farmers. The Rural Electrification Administration pressed for the extension of central-station service in the farm areas.

Marketing agreements were made for many different commodities under the appropriate sections of the original AAA. The milk marketing programs are interesting examples in this field, in that action was taken by both federal and state agencies. Under these programs milk classifications according to use (fluid milk, cream, butter, and other manufacturing uses) were developed for the major milk markets, and prices were established and enforced for these classifications. State programs were limited primarily to the intrastate markets, and milk control boards and commissions were organized in most of the states. These parallel developments have sometimes involved conflicts—conflicts emphasized by the broader issue of "states' rights"—both as to jurisdiction and objectives, but there has been enough coördination and coöperation to suggest the possibility of joint state and federal action in such fields.

Farm program of the Second World War: Production and prices. Agricultural developments in the Second World War differed in many respects from those following 1914. As the diagram given in Figure 40 on the following page indicates, production increases were more pronounced, and this expansion was more generally spread over the whole field of agriculture. Wheat and hog numbers again increased, but eggs, poultry, fats and oils, dairy products, and many others also felt the stimulating effects of wartime demands.

As shown in Figure 40, the development of agricultural prices was essentially similar in both of the war periods, despite some attempts at control. Between 1940 and 1942 the over-all effect of the increases in farm prices was to offset the higher level of prices paid by the farmers. After 1942, however, the farmers gained an advantage comparable to that of the First World War. Gross farm income increased from 11 billion dollars in 1940 to 23.4 billion dollars in 1944, while national income was increasing from 78.5 billion dollars to 160.7 billion dollars. The income-parity ratio increased rapidly from 90 in 1940 to 150 in 1943 and then dropped slightly to 145 in 1944.

One result of farm prosperity was again a rapid increase in land

values. In March, 1944, land values were 38 per cent higher than before the war as compared to a 40 per cent increase for the comparable period in the earlier war. In the first period the inflation of land values brought with it an expansion in the farm mortgage debt, but farmers seemed to have profited from that experience and used their improved income in the later period to reduce rather than increase debt. Farm mortgages were reduced by more than 1.3 billion dollars between 1940 and 1944, when they stood at the lowest level since 1916.

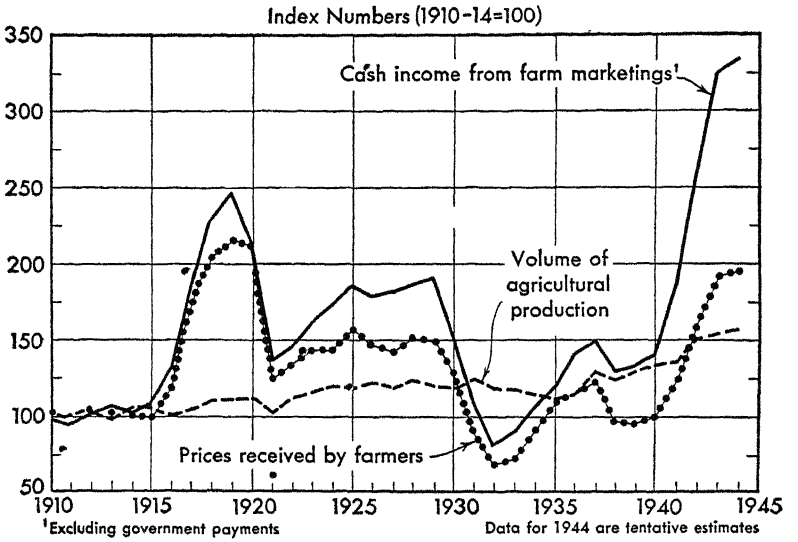


FIG. 40. FARM PRODUCTION AND PRICES IN THE UNITED STATES, 1910-44 [1]

Price "floors" and "ceilings." The Agricultural Adjustment Act of 1938 is still in effect, though it has been changed significantly by a series of amendments culminating in the Steagall amendment of 1941. This amendment replaced the 52 to 75 per cent of parity loans on the basic crops (cotton, wheat, corn, rice, and tobacco) by loans at 85 per cent of parity, and provided for the support of any nonbasic commodity whose production needed to be expanded at 85 per cent of parity or of a "comparable price." Moreover the

[1] Figures taken from Bureau of Agricultural Economics, *Agricultural Statistics*. Washington: Government Printing Office, annual report.

period for which these price supports were to apply was not to be terminated before the Secretary of Agriculture had given sufficient public announcement to permit producers to make readjustments in production.

In October, 1942, Congress amended the Steagall amendment by raising the minimum level of price floors to not less than 90 per cent of parity or comparable price, save for cotton which was to be supported at 92½ per cent of parity, and by adding peanuts to the list of basic crops. In addition the support period was to continue "until the expiration of the two-year period beginning with the 1st day of January immediately following the date upon which the President by proclamation or the Congress by concurrent resolution declares that hostilities in the present war have terminated." Since 1942 a number of further revisions and price supports have been announced.

The primary purpose of these price floors was to stimulate production by guaranteeing the prices that farmers would receive and by assuring them that these prices would hold for some time in the future. This was intended to make it possible for farmers to avoid some of the uncertainties that had always characterized future prices and so to plan a more rational basis. To further this purpose the Secretary of Agriculture announced, soon after Pearl Harbor, a comprehensive list of farm production goals. All of these price-supporting programs were carried out through purchases of commodities for military, lend-lease, the school lunch program, or other governmental uses and programs, or through loans, purchases, and other operations conducted by the Commodity Credit Corporation.

Along with this series of legislative acts supporting agricultural prices, another type of legislation came to the fore. This was symbolized by the Emergency Price Control Act of 1942, aimed at keeping prices from rising by providing ceilings to check "speculative price rises, price dislocations, and inflationary tendencies."

Tied closely to the price control legislation was the payment of governmental subsidies to the producers or sellers of certain farm products. This program was based on an executive order issued after the signing of the so-called Anti-Inflation Bill of October, 1942. It authorized direct or indirect subsidies through purchase at high prices and resale at lower prices where needed (1) to insure the necessary production or distribution of any commodity, (2) to main-

tain ceiling prices, or (3) to prevent a price rise. Among agricultural commodities on which such subsidy payments were important are fluid milk, butter, cheese, soybeans, flaxseed, peanuts, wheat for feed and industrial alcohol, and dry beans and peas.

The stresses and strains of the war period served to emphasize the conflict between policy objectives. Thus the two fundamental lines of price legislation were potentially at war with each other. While one was concerned with increasing agricultural prices, the other was charged with preventing price increases. As a matter of fact, ceiling prices were the more important during the war. Despite the fact that in 1944 there were price-support programs in effect for ninety-one individual commodities, most farm products were selling above their support prices. Volume of production was the real goal during the war, and not the raising of prices. To achieve this in the face of ceiling prices the incentives of subsidy payments and of announced production goals were considered essential.

Agricultural production during the war. The product of the American farms actually rose, during the war period, far above prewar levels. For example, in the seven "fat years" of 1940–46, the harvest of corn, America's most important feed grain, was 21 billion bushels, whereas in the seven previous lean years only 15.4 billion bushels were harvested.

This remarkable record was the result of a combination of causes. The government's program doubtless contributed. To some degree the result reflects an increase in acreage under cultivation. But more important by far was the amazingly good weather with which the farmers were blessed. Other developments contributing to these high yields were the increased mechanization of farm tasks, the use of better cultural practices and improved varieties of seeds, and the ability and willingness to work hard for long hours. Working thus, with worn equipment and inadequate supplies of labor, the American farmers performed one of the outstanding production feats of the war.

Appraisal of farm policy: Guiding principles. Economists are united in pointing out that a system of free and competitive market prices will bring an optimum allocation of resources among the various alternative uses.

Applying this principle to agriculture, we observe that the operation of demand and supply is accompanied by lags, frictions, and

especially lack of knowledge on the part of the farmers. To a considerable extent these reflect the biologic nature of the processes of agriculture. Moreover it appears that farmers' decisions are, to some extent, made on noneconomic grounds. These factors complicate the situation and lead to special problems of agriculture, as we have seen.

Nevertheless the fact remains that agriculture represents as close an approximation to pure competition as can be found in our economy. Large numbers of relatively small producers are engaged in the production of relatively uniform and standardized products.

Efforts to improve the well-being of the farmer class by governmental policies interfering with the free play of the competitive market raise questions of basic principle. Not only farm people, but also those living in cities and towns, are affected by these welfare considerations. When society attempts to provide a minimum living standard for any of its members, it interferes with the distribution of incomes on a strict productivity basis. This would be true even if the general level of productivity were high and relatively stable. From the economic standpoint, the more any program affects the use of resources, the more it will interfere with the efficient conduct of production. In short, the least disturbing program of distribution of the product is that which permits the allocation of resources to follow the competitive pattern.

Parity prices and income. These fundamental principles have not yet received sufficient attention in the development of agricultural policy. The ultimate objective of the farm program has been to increase farm income to that point where the net income per person on the farm would bear the same ratio to the net income per person not on the farm as it did in the period 1909–14. This concept of parity income obviously appeals to its beneficiaries. It may also appeal to certain others as a fair division of the national income—as according simple justice to the farmers. To such persons the method of reaching this goal, involving the raising of the prices of basic commodities to their parities through restrictive devices and government purchases and loans, may seem simple and easily understood, and its results may appear to be readily measured.

However, fixed prices—and parity prices are in essence fixed prices, even though they may shift with other prices—are incompatible with a system of free enterprise. In such a system competi-

tive price reflects changes in demand and supply and is relied upon to direct production into channels most profitable to society. To tie the prices of the basic commodities to a standard determined by conditions in the past means abandonment of price as a regulating force.

The situation in 1933 called for elimination of submarginal and marginal land held in cultivation in the hope of better times, but percentage reductions in acreage devoted to the basic crops gave no such incentive, for benefits were paid to all alike. As a result, production was maintained in part on inferior soil and reduced in part on better land, with resulting loss in social income.

Nor did the program solve the problem of surplus commodities, since despite restrictive measures, total agricultural production was not reduced. To give but one example, the average acreage planted to wheat in the period 1940–42 was lower by 13 per cent than in the years 1931–33, but the production of wheat was higher by 21 per cent.

The Commodity Credit Corporation found that, if prices were to be maintained, it would be necessary to remove the surplus from the open market by purchase and by commodity loans. In the early 1940's the Corporation owned or had pledged to it as security for loans, 10.7 million bales of cotton, 519 million bushels of wheat, 477 million bushels of corn, and quantities of tobacco, rye, and barley. From 1933 to the end of 1941 the total amount of purchases and commodity loans amounted to almost 3 billion dollars. What the result would have been had war demand not intervened no one knows, but the situation was critical, and it is no exaggeration to say, as does one authority, that "the War rescued the CCC, by converting into valuable assets (except for cotton) the huge inventories that the CCC was instrumental in accumulating." [1]

Conclusions. The United States Government was presented in 1933 with an acute agricultural emergency. From both the political and the humanitarian points of view, it seemed to those in authority that something had to be done for the farmer, and done at once. It is not our purpose here to criticize the resulting legislation, which was frankly of an emergency character.

The prospect of the development of this program into a permanent policy, however, is another matter. There are strong political

[1] Schultz, T. W., *Agriculture in an Unstable Economy*. New York: McGraw-Hill Book Company, 1946, p. 175.

forces which oppose abandoning the present farm program without finding a substitute that will give the farmers greater economic security than they have in the past. Not only are certain farm organizations and some politicians committed to programs of agricultural production and income adjustment, but there are doubtless others who feel that the major insecurities in agriculture come from causes beyond its control and to whom it seems natural and proper for society, as a permanent policy, to assume some of the risks.

The weight of economic principle is against this program. The payment of millions of dollars in agrarian subsidies, the loss of export markets through artificially high prices, the possibilities inherent in government crop loans—these and other phases of the program are serious matters. The theory of parity prices is not only of doubtful validity itself, but its practical application is attended with many dangers. The tying of agricultural prices and incomes to some historical period cannot fail to stifle those economic adjustments that are essential to the upward trend of efficiency and productivity. Prices are not goals, but only the methods by which a free economy allocates resources among the myriad of possible alternatives, to the end of maximizing production for the satisfaction of human wants.

So long as agriculture is not wholly competitive, the public interest will be served by much more dependence on the free action of the market and by such governmental programs (if possible) as would tend to bring prices closer in line with the competitive model than to freeze them in a pattern that has long been obsolete. The problem of the allocation of resources between agriculture and the rest of the economy is more complicated, for the noncompetitive elements in many sections of industry are more important and persistent. But here again the public interest lies in the direction of competition in a market free from government restrictions or control.

CHAPTER XXXVI

Transportation

Introduction. Economics of transportation. The economic organization to create place utility by transportation has many of the same characteristics found in organization for other forms of production. The importance of the technological advance over the years has been as great as in manufacture or agriculture. The development of improved economic institutions to facilitate a constant betterment in service has been equally significant, and at times transportation has pioneered in the creation of these institutions. It was the proving ground for the corporation, and for government regulation by an administrative agency.

From the point of view of economic theory, the allocation of resources to transportation enterprises is not unlike that in other fields. Competition, monopoly, and the various intermediate patterns are all to be found. Economic rent, as a result of location and monopoly power, is an evident phenomenon. The fact that transportation is so much a concern of all has meant that data regarding prices and costs and other economic factors have been more available than in other enterprises. Thus there is unusual opportunity to see basic principles in operation and to see the applicability of economic analysis.

Topographical factors: General. The aspect of transportation activity that is unique relates to the geographical factors, which leave their imprint regardless of the institutions surrounding transportation or of technological advances in the industry. Of these factors the topographic is the most fundamental. The nature of the coast line designates the ports for ocean shipping and plays a part in the safety of navigation. The distribution of navigable rivers and lakes determines the possible channels of inland water transportation and determines also the possible obstructions to the path of highways and railways. The location of water resources is critical for canals. Hills and mountains are obstacles to land transportation, and passes through them are the funnels through which commerce

must be fed. Climate decrees what portion of the year waterways are free of ice, how forbidding is the threat of flood, and the degree to which storms and fog may be an obstacle on land and in the air.

Where a variation in any of these characteristics forces a change from one type of transportation to another, the work of transshipment is involved. This locates commercial activity such as warehousing and brokerage, with their influence on other aspects of economic activity.

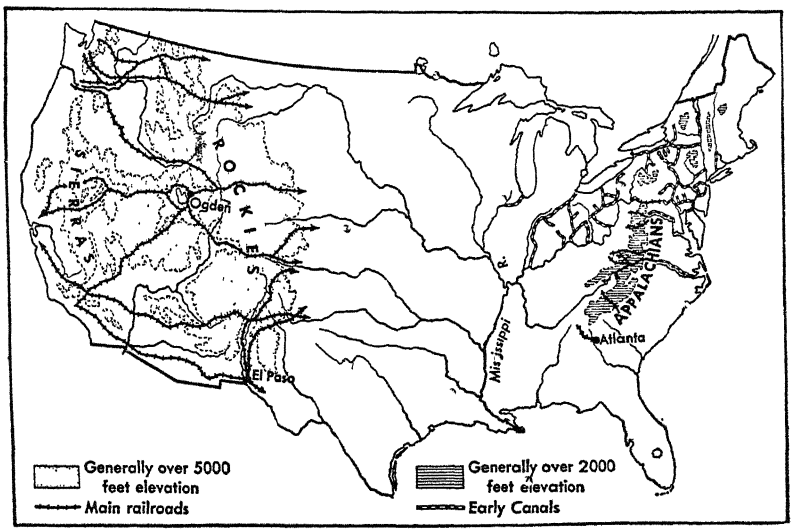


FIG. 41. TOPOGRAPHY AND THE DEVELOPMENT OF TRANSPORTATION IN THE UNITED STATES

The impact of all these factors was particularly important at the earlier level of technological development. Even today railroads cross big rivers at relatively few points and large mountain ranges through but a few passes. Water transportation fails to reach vast areas of the world. Air carriers find that the usability of airways varies, depending on the extent to which they are fog-bound. Population and commerce still center at the points of transshipment between different types of carriers.

Topography and the development of the American transportation system: Early development. The importance of these topographic factors may be readily seen in the development of the

transportation system of the United States (Figure 41). The earliest important transportation in this country was by natural waterway with wind and animals supplying the motive power, and on land by animal power. Natural water transport was by far the cheaper and thus commanded the long haul trade in low-value commodities. But being restricted to the ocean and a relatively few deep and slowly flowing rivers, water transport was available to but a limited area of the country. Land carriage, on the other hand, was so expensive that low-value materials could be moved only locally, while long-haul transportation was limited to a thin stream of high-priced goods or passengers.

The cheapness and great load-carrying capacity of water transportation provided a strong incentive to develop man-made waterways to supply the regions where natural waterways were not available. During the early part of the last century, both private corporations and state governments engaged in extensive canal building throughout the Northeast and Middle West.

Though the canals could provide water transportation for a much greater area of the country than navigable rivers, geographical characteristics still limited their coverage. Canals require a large supply of water for lock operation, and high hills make for an impossible number of locks. Nevertheless, by the 1840's, three thousand miles of canals had been built, and another three thousand miles were seriously projected. But now the time was late, for another technological advance developed, offering many of the advantages of the canal while excelling as a year-round carrier in northern climates. This was the railroad.

Topography and the railroads. The railroad was first viewed merely as an improvement on the turnpike, with the expectation that it would offer more expeditious service than the stage coach for passenger, mail, and light-weight transportation. Due to the coefficient of friction between iron rail and iron wheel, the gradient allowable on the ordinary railway was much less than was possible on the highway. Railways were thus confined to terrain less hilly and rugged than turnpikes. Moreover the great weight of railway vehicles makes the crossing of waterways more difficult.

It was ten or fifteen years before the railroads proved that they could take the freight business from the water carriers, but when this had been demonstrated, the railroads spread rapidly into the region west of the Appalachians, and by combining short lines into

larger systems, they joined the Atlantic Coast with Chicago, and the Mississippi and the Mid-west with the Pacific and the Southwest. By 1910 the consolidation process had proceeded to the point where 96 per cent of the rail business of the country had been concentrated in about 150 companies, each having annual revenues exceeding $1,000,000.

This unfolding of the country's railway system, whether in the East or the West, was conditioned by physical characteristics of the respective parts of the country. In the East the contours of the Appalachians determined where railroads could be built to carry efficiently the great volume of commerce with the Middle West. The favorable character of the Albany-Buffalo route, which made the Erie Canal so successful, provided the first through rail connection to the West in 1842 and an additional right of way for the West Shore in 1888. The mountains of Pennsylvania, prohibitive for canals, were crossed by one main rail line in 1852, and several secondary ones later. Not until 1853 did the Baltimore and Ohio reach Wheeling. Twenty years more were needed before the Chesapeake and Ohio penetrated to the Ohio River. The first main-line crossing went west from Ashville to Atlanta, Georgia, in 1882, and the only other in 1900.

Just as the relatively level country of middle and western Ohio, Indiana and Illinois provided fertile ground for canal building, so it made possible any number of rail lines. By the time of the western railroad era the population had moved farther west than during the canal era, and so the equally favorable profiles of Iowa, Wisconsin, and Minnesota, as well as of the tier of states just west of the Missouri, provided a railroad builders' paradise.

The other geographic element in the transportation pattern of the Middle West was the presence of three great rivers, the Ohio, the Mississippi, and the Missouri. They continued as water routes with steamboats competitive to rail operation, and in addition they were wide enough to make bridging by the railroads difficult and expensive. The first bridge crossing of the middle Mississippi was achieved in 1855, but it was not until 1936 that the opening of the Huey Long bridge at New Orleans provided the first crossing below Memphis.

Crossings of the Missouri and the Ohio were so expensive as to be economically possible at only a limited number of points, and the commerce across those rivers has always been funneled through a

limited number of "gateways," such as Cincinnati, Louisville, Henderson, Metropolis, and Cairo on the lower Ohio. This has given particular commercial importance to these cities as compared to other places along those rivers.

West of the broad Mississippi Valley, it is the great ranges of the Rockies, Sierras, and Cascades that have molded the railroad pattern. Feasible passes are few, and all but the Southern Pacific's southern line are forced to a height of more than 5,000 feet in crossing the Rockies within the United States. In all there have been only seven rail crossings.

The two western ranges can be crossed more easily; in their north and south sections there are several available passes no higher than 3,000 feet, with the Columbia River providing a unique river route. But in the middle, the Southern Pacific's "Overland Route" must cross at 7,000 feet in the historic Emigrant Gap route. The location of each of these trans-mountain routes has been a controlling factor in the location of the economic centers of the country's mountain regions.

These are some of the major topographic factors that have affected the geographic pattern of the United States railways. The Great Lakes too have left their imprint; so has the St. Lawrence River. To a lesser degree, the pattern has been influenced by smaller mountain ranges such as the Ozarks, smaller rivers such as Hudson or Rio Grande, and estuaries or bays such as the Chesapeake. These major and minor factors determine the lines along which railways are built; they also affect the way rates are made and the speed and other qualities of the service that is offered. As a consequence they profoundly influence the specific location of much of the country's economic activity.

Topography and the modern highway system. The modern highway is influenced far less by topographic factors than is the railway. For each railway crossing of the Appalachians, there are several main highway crossings, as well as many secondary ones. In generally rugged country, where railway construction may be expensive, the highway finds far greater opportunities. The railway engineer, even in the most mountainous territory, attempts to keep to a grade of 2.2 per cent or a 116-foot rise per mile on the worst rises. Such a slope is scarcely noticeable on a road. Curvature limitations, too, are more liberal for a road.

Compared to the railroad, the highway can go almost anywhere.

This is illustrated by the fact that the country's rural highway system has some 3,000,000 route miles, while the maximum reached by the railroads was 250,000 miles.

The highway, along with the motor vehicle, can be credited with making available more locations for light industry and commerce than did the railroad, and with allowing a dispersion of location within key regions determined by other factors.

Transportation Costs: General. One of the critical problems in any form of economic activity is the way in which the costs of producing goods or services are incurred. Costs in the transportation field have been subjected to more critical analysis than those in other areas. This is partly because of the publicity given to railroad rates and the general expenses of transportation service; it is partly because of the marked difference in the prices charged for what seem to be similar services. Prices will be discussed later; here some of the bare essentials of the cost problems will be analyzed.

Fixed costs. In the days when railroads were expanding and new facilities were being built, the capital costs of interest and rentals took a very substantial portion of the revenues. At the start facilities did not secure the volume of traffic that they were designed to carry in the long run. The designs may have been extravagant, in the sense that they called for greater capital expenditure than prospective revenue in the immediate future would warrant. The result was that interest charges appeared as a large fixed cost, which had to be met by whatever volume of business could be attracted.

It also appeared that certain of the operating expenses went on, regardless of volume of traffic. For instance, certain repairs to track and roadway cost little more for five trains a day than for two.

Such capital charges, together with the repair expenses, constituted the so-called fixed costs. They were the costs which could not be varied directly with the amount of business or the revenue taken in. In 1900, taking the railroads of the country as a whole, statistics showed that fixed interest and rentals equaled about 20 per cent of the revenue of operating companies. Repair costs of track and way called for another 14 per cent, a substantial proportion of which was considered to be fixed.

However as the railroad system has become mature, the character of the traffic demands has become known. The amount invested in facilities has become closely adjusted to the volume of traffic for all but branch line types of operation. This is in part

because the heavily traveled lines are built to the highest standards and make use of the latest technological advances. It is also because a permanent management regulates its capital expenditures in accordance with traffic and revenues, whereas the promotional builders of the earlier days gave little thought to this relationship. Since the fixed interest charges are a function of investment, all this means that they are more closely related to the volume of traffic than was initially the case.

Furthermore, over the years, an increasing share of revenue has gone to labor rather than to capital. In 1900 wages were equivalent to but 39 per cent of revenue; in the half year ending June, 1945, they were equivalent to 52 per cent. In turn fixed interest and rentals were roughly 20 per cent in 1900 and only 7½ per cent in the latter period.

Variable Costs: General. It has been found that modern administrative control keeps expenses from year to year pretty much in line with revenue, which in the short run is an index of volume of business. Over the long run, the ratio of expenses to revenue varies with changes in technology, wage requirements, and material costs, but in the short run, these are not controlling. From 1900 to 1910, the ratio of expenses to revenue did not vary more than 3 per cent, in spite of the fact that revenue increased from $1,500,000,000 to $2,800,000,000. Again from 1925 (after the effects of the First World War had worn off) to 1940, there was no greater variation in the ratio. This was in spite of variations in revenue from $6,500,000,000 to $3,100,000,000, and back to $4,400,000,000.

This does not mean that on a lightly traveled branch line business cannot be increased with a less than proportional expenditure, but it does suggest that on the densely traveled lines changes in volume of business from year to year bring about nearly proportionate changes in expenses. The lion's share of revenue and expense comes from the latter type of line, so that the total picture seems to indicate a substantial flexibility of expense under modern conditions.

The implications of this as to pricing are that, unless the added business is on branch lines alone or only for very short periods, nearly full costs must be considered in rate making if profits are to be preserved. In theoretical terms, this implies that under most conditions the difference between marginal cost and average cost for regular flows of traffic is not very great.

Utilization of vehicle-carrying capacity and variable costs. What has just been said applies to the over-all costs in different broad categories. When we get down to the detailed unit of operation, the car, we find that the degree to which a car can be filled to capacity has a great influence on costs per unit of service. A freight car can carry but four Ford, Chevrolet, or Plymouth passenger automobiles, which means a load of only 6 tons; the same car can carry 50 tons of copper ingots. The cost of handling the car does not vary greatly between the two loads, but if the cost is figured on a ton-mile unit basis, the costs are eight times as much for the automobiles as for the copper.

Coal, which is loaded to 70 and even 100 tons per car, provides an even greater contrast. The apparently low rates on coal per ton-mile are really not low, but may even be high, because coal makes such good use of the cars' carrying capacity. The high earning power of those roads which haul coal almost exclusively are evidence of the way that good utilization of capacity affects costs and consequently earnings.

A passenger coach which may haul on the average twenty-five passengers in ordinary times costs no more to handle than a sleeping car hauling on the average only eight passengers. In general the cost per passenger-mile is therefore in the ratio of 1 to 3.

In wartime, when people change their travel habits and are willing to accept less in the way of comfort and privileges, the average load of the coach may rise to fifty. There is no appreciable increase in costs. The capacity utilization or load factor is increased twofold; the unit cost declines 50 per cent, and if rates remain the same, the profits increase correspondingly.

It was principally the ability to utilize more efficiently the capacity of both freight and passenger facilities that enabled the railroads to carry the tremendous volume of traffic during the Second World War. For the same reason ton-mile and passenger-mile costs went down and earnings up. The reversion to the degree of utilization that is acceptable in peacetime in turn reduces the railroad capacity. Other things remaining equal, unit costs increase, and profits decrease.

The ability to fill up the basic unit is all-important, whether it be the car of the railroad, the ship of the water carrier, or the plane of the air lines. In 1940 airplanes were flying with less than 60 per cent of their seats occupied. With wartime conditions the utiliza-

tion rose to nearly 90 per cent: costs per passenger-mile dropped. But with the inevitable reduction of the load factor to normal peacetime standards, unit costs have risen.

The implication of this for pricing is clearly that any price pattern which gives an incentive toward better utilization will yield important returns in profit. For this reason, in part, rates per 100 pounds for carload lots of commodities are lower than such rates for less-than-carload lots. More recently there has been a tendency to introduce rates which are lower per unit, if a larger number of units can be loaded in a car. Truckers have adopted similar differentials in their rates.

Competition between carriers. Development of competition. Two basic characteristics have done much to set the competitive economic pattern of the American transportation market. The first, arising out of the almost continuous technological advances, is that, except for the earliest years of its history, the transportation system has been made up of one type of transportation superimposed upon another or others already operating. The canals came after the elementary roads; the steam railroads after the canals, steamboats, and turnpikes. The electric street and interurban railways followed the railroads. Then the motor vehicle entered the scene, after those two types of rail transportation were well established. Finally the air carrier appeared.

The second characteristic is the fact that geographically and economically separate organizations have built up the facilities for each type of transportation. The national government did not build the early canal system as a whole; rather the several states and small private companies individually built up separate sections that were to make the whole.

Likewise with the railroads, literally thousands of separate corporations built lines in one part of the country or another. Some were strictly local lines, others were long, joining one part of the country with another. Together they all provided the network over which the freight and passenger cars of the country could move, once the gauge was standardized in the 1880's. In all there were some 150 principal separate companies, each with its own particular coverage of the country.

The consequence of these two characteristics has been that, as the years have gone on, there has been no basic physical restriction on competition in a large share of the transportation market. Since the

Civil War, for instance, there have always been several separate through lines of railroad and many combinations of shorter lines providing service between New York and Chicago. There have also been a number of steamboat lines on the Great Lakes, which have connected with Erie Canal carriers or with eastern railroads. More recently there have been added to this array hundreds of individual truck carriers and a number of air lines. What may be said for the New York-Chicago trade route applies equally to other heavy-volume channels of commerce. In earlier days, only the more sparsely populated or topographically handicapped points were limited to single-carrier service. More recently the motor vehicle and airplane have provided competition even at these points.

Restriction of the number of competing carriers. Confronted as they were with aggressive competition among themselves, existing carriers have continuously endeavored to prevent or restrict new competition. Originally they tried to have their charters so drawn by state legislatures that no other similiar companies could operate in their territory. When this type of obstruction was rendered impossible by the enactment of general incorporation laws, they turned to the acquisition of control over land or river crossings essential for the establishment of competitors.

With the coming of governmental regulation of transportation, the pattern for restricting transportation enterprises has in the main taken a different form. While such regulation was originally intended for the protection of the public against discriminatory rates and service, the regulatory powers have with time expanded to include control over entry into one phase of the transportation business after another.

The Interstate Commerce Commission, established in 1887 as the federal regulatory agency for railroads, was given the power in 1920 to prohibit railroad construction unless it was shown to be necessary and in the public interest. Coming as it did, almost a hundred years after the first railroads of the country, and after the railroad network had been well developed extensively and intensively, this control was not of wide significance in determining added competition.

However the pattern of the railroad field set a precedent for imposing government certification for entry into the motor carrier industry in its early development. Control was adopted first by states and then, in 1938, by the federal government. The previously es-

tablished carrier, the railroad, was not the least powerful in the advocacy of this restriction. In the motor carrier industry the power of refusing the right to enter markets has been used aggressively. As a result, the number and extent of operations of firms in the transportation market have been markedly less than they would have been had the entrepreneurs' interest in profit been the sole factor in determining whether a new operation should be initiated.

Existing carriers by 1940 succeeded in getting Interstate Commerce Commission control over the entry of common carriers into domestic water transportation. In 1938 common carrier air operations had been made subject to similar restrictions applied by a separate regulatory agency. Thus by 1940 every firm desiring to enter into any area of the general transportation market in interstate commerce, except for petroleum pipe lines, had to obtain a certificate to do so from some federal regulatory agency. The certificate to enter the market was granted only if it could be demonstrated that there was not already sufficient service. The new entrant usually had to argue his case against a strong presentation by existing carriers.

This growth of regulation can be attributed mainly to the desire of existing carriers to avoid added competition. The results have not been too unsatisfactory to them. But no longer can it be said that, if the prices and profits of a transportation operation become abnormally high, new entrants (whether operating the same or a new type of transportation) will automatically be brought into the field to reduce such prices and profits.

Competition in rates: Conditioning factors. The combination of technically different forms of transportation and geographically differentiated and independent carrier firms has left a deep imprint on the price pattern. These factors were particularly important when that pattern was in its formative stage, between the Civil War and 1910. In this period two technically different forms of transportation were developed—the railroad and the steamboat.

As has been indicated, the topography of the country restricted the extensive development of the latter. There was a limited number of places along the coast where vessels could find harbors, while the usable inland waterways were confined to the larger rivers and the Great Lakes. This meant that only few cities could be served by both rail and water. In the case of the railroads, there was the opportunity for a dense network of lines in the easy terrain of the

Middle West. On the other hand, in the hilly and mountainous areas of the country, the opportunity for several railroads to gain access to a city site was limited.

Taking these two circumstances into account, shipping points could be classified as of three types. There were those having both water and rail carriers, those having several rail carriers, and those with but one rail carrier. In general, the existence of competition between the two types of transportation made for low rates at the port, whether river or ocean. Where several independent rail carriers tapped the same point, rates might also be low, but points served by a single railroad and no water carrier tended to have high rates.

Inequalities due to competition. A typical illustration is provided by the rates on a particular class of goods in 1910, from St. Louis down the Mississippi Valley route to New Orleans. For the 322 miles from St. Louis to Memphis, a river port, the rate was 65 cents per 100 pounds. But the rate to Lula, Mississippi, 47 miles further south, without water or alternative railroads, was almost twice that to Memphis—$1.17. For the haul 103 miles beyond Lula, to a second river port, Greenville, Mississippi, the rate was only 90 cents. To an inland, noncompetitive rail point, 36 miles beyond Greenville, the rate was $1.26. Adding 14 miles more to the haul, to another inland point, Jackson, which had more than one railroad, the rate was only 98 cents. Under the transportation handicap thus imposed on single-carrier points, it is not surprising that in none of them had the population grown to over 2,500, whereas the lower rate sites all had populations of over 10,000.

A second illustation is provided by the rates effective at about the same time on the central transcontinental route. From the port of New York to the port of San Francisco, the rate on a certain class of goods was $3.00. But for a haul shorter by 246 miles—from New York to Reno, an inland point with one railroad—the corresponding rate was $4.29.

By contrast, in an area like Indiana, where water transportation to the East was not significant and railroads were both plentiful and cheap to build, there quickly developed a rate scale, uniformly increasing with distance, applicable for all points to and from New York. For the sake of simplicity, as well as to equalize commercial opportunity for near-by places, individual rates were not quoted

for each station, but small districts were defined, and all points within those districts took the same rate.

Competitive equalization of rates. While the pressure of competition could cause inequalities in rates for equivalent distances, it could also equalize rates for considerable variations in length of haul. In 1879 the Louisville and Nashville was the only carrier leading out of Cincinnati and Louisville to Atlanta. The rate on certain items for the 475 miles from Cincinnati to Atlanta was $1.39 and on the same items from Louisville to Atlanta, 451 miles, it was $1.19. A year later, after a second railroad was built south from Cincinnati to independent connections with Atlanta, the $1.39 rate dropped to $1.19. With competing railroads, the two Ohio River cities were given equal opportunity ratewise in Atlanta in spite of the difference in distance. Had there always remained a single railroad having sole control of the service south of the Ohio, Cincinnati would always have had to pay a price for its 24-mile greater distance to the southern market.

Geography and rate competition. The varying geographic coverage of the separate railroads, together with alternate sources of production within the country, has developed competitive forces that account for other aspects of the rate pattern. As a result of the way railroad companies have grown up in this country, we find that no company serving the North Atlantic states extends west of the Mississippi cities or south of Virginia. None serving the Pacific Coast or Rocky Mountain area reaches east of the Chicago area or the southern Mississippi cities. This is in contrast with the possibility that the country might have been served by several competing roads, each covering the whole country. In turn particular commodities are usually produced in several parts of the country and are sent to numerous destinations. The relation between the location of these origins and destinations and the geographical pattern of individual railroads has a strong influence on the making of rates.

The relation of certain key rates on orange traffic illustrates how the actual arrangement of roads, in conjunction with alternative sources of production, molds the rate pattern. The production of oranges started in Florida. To reach the big northern markets, of which New York and Chicago may be taken as representative, the railroads of the South joined with those of the East to make through rates. In 1907 it cost 75.5 cents to send a box of oranges from a typical shipping point in Florida to New York, a distance

of 1,242 miles. The presence of coastwise water carriers operating from Florida ports to New York had its effect in keeping the rail rates, on the railroads parallel to the coast, lower than they might otherwise have been.

As the California orange production developed, the western railroads were interested in stimulating its increase, in order to make a profit out of hauling oranges to the big eastern markets. The rates could not be too far out of line with rates from Florida, although the difference in seasons between Florida and California provided a degree of insulation with respect to producer competition. But from California it was 2,265 miles to Chicago and 3,149 to New York, instead of the 1,200- to 1,400-mile distance separating these cities from the South.

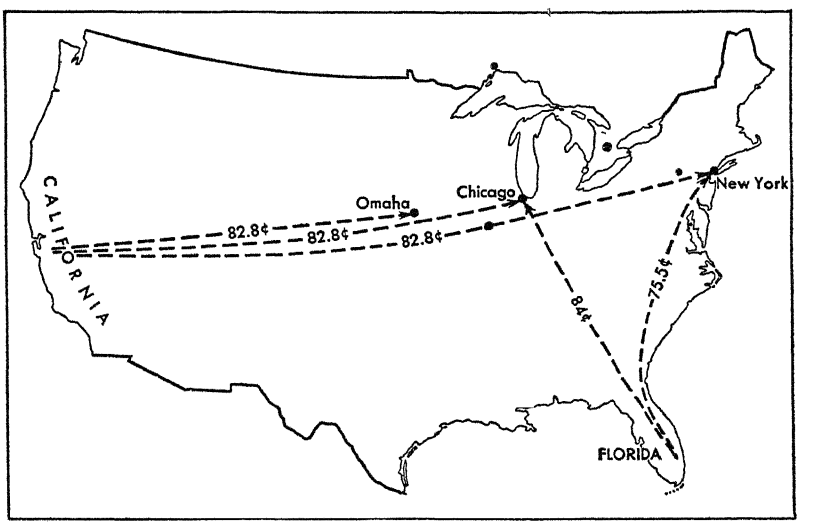


FIG. 42. FREIGHT RATES ON A BOX OF ORANGES FROM CALIFORNIA AND FLORIDA TO CERTAIN CITIES, 1907

Nevertheless the western roads set a rate to Chicago of only 82.8 cents a box for a slightly smaller box than the Florida one. Those railroads met Florida competition in Chicago. By agreement with the eastern roads, the rate of 82.8 cents was made to apply also to the New York terminal 900-odd miles farther on. This did not meet the southern competition as well in New York as in Chicago but it was better than if something were added to the price for

the extra mileage. For nearer towns such as Omaha and Denver, the western roads still charged a rate of 82.8 cents in spite of the shorter haul, because the southern roads by themselves could not make competitive rates to those points.

Thus as shown in Figure 42, we have over half the country being given the same rate from California on oranges, with the difference of 2,000 miles of hauling disregarded in the price of transportation service. Again had carriers, each having country-wide coverage, been making these rates, distance would not have been disregarded in this fashion.

Conclusions. It can be seen that geographic and technical factors have made for an irregular impact of competition on prices for transportation service. The charge that railroads are naturally monopolistic can hardly be supported in a country-wide sense, though it did apply for many small towns or in topographically difficult areas. Clearly, in those cases where a single carrier felt that it had control over shipping from a particular point which could not be easily challenged, that carrier acted monopolistically. Because of this, business preferred (opportunity allowing and other things being equal) to choose multiple-carrier points for location, and as a result, industrial and commercial population has tended to concentrate at those points. The consequence is that the most industry and traffic is carried on under conditions where competition, unless artificially restricted, is naturally present.

Specific rate determination: Bargaining of shippers. During the period when the rate pattern was being established, an important nongeographic force in shaping rates was the stronger bargaining power of the large shippers compared to that of the small ones. Aside from any economies of handling large-volume shipments, the bigger shipper was able to bring pressure to bear on the railroads for the lowest possible rates. Such a shipper was frequently located at multiple-carrier points, where he could play one railroad off against another to get preferred rates. Losing his business might mean significant losses to any one carrier. In addition such a shipper often had connections with carrier managements or boards of directors, which could be used to obtain preferential treatment. By contrast, an individual who made a single shipment in a year, or a small businessman whose share of the traffic might be a "drop in the bucket" compared to the total, could bring little bargaining power to bear on a rate issue.

"Class rates." One result of this was the creation of two types of pricing for freight service. One type was for the freight of infinite variety coming from the many relatively small shippers. Every kind of freight was listed in a class, in respect to which class a particular rate could be quoted between any two points. Technically the rates used in this connection were known as "class rates.

In the early days each railroad set up its own classification and its own class rates, but by the turn of the century there was unification classwise within each of three broad regions. These not unnaturally fell into the geographic pattern of the railroads already mentioned. There was the eastern or "Official" region, including the railroads north of Virginia and the Ohio and east of the Mississippi and Chicago. The "Southern" region extended south from the "Official," and the "Western" region covered the area to the west of the other two.

In general items with great bulk for their weight, those with high value, or those with a tendency to break, were put in the classes taking the higher rates. Dense, low-value, and hard-to-damage articles were put in those taking lower rates. To the extent that value was a determinant of the class, there was clearly a monopolistic factor involved in the market for which these rates were designed.

The class rates were made with reference to smaller regions than the classifications. The "Eastern" area was divided into three subregions: New England; "Trunk Line," west of that to a line through Buffalo and Pittsburgh; and "Central" for the rest. The "West" had a "Southwestern" subdivision applying south of Kansas and mid-Missouri and east of the Rockies, a "Western Trunk Line," north of there and east of the Rockies, and several further subdivisions in the area west from the Rockies. In the South there were not such marked subgroups. (See Figure 43, page 761.)

"Commodity rates." The second type of prices quoted for freight transportation were the so-called "commodity rates." These usually applied to carload lots, though they were not uncommon for smaller shipments. The rates were established separately for each of the many particular commodities which moved in substantial volume over regular channels. Coal, lumber, wool, ammunition, cotton piece goods, etc., moved under such commodity rates.

Commodity rates were lower than the class rates for similar service. The orange rates previously discussed were of this type. There

was no classification of commodities, because each was given particular rates over the routes along which it was usually offered for transportation.

With this individual treatment, it was possible to adjust rates to the varying patterns of production and market location, as well as of the carrier competition illustrated above. Thus each commodity had its own unique rate structure. These over-all structures knew no regional boundaries, but covered as much of the country as was affected by the regular flow of the particular commodities. In practice an increasing proportion of the traffic has been carried under commodity or similar rates, so that currently they apply to roughly nine-tenths of railroad freight.

Rate discrimination. The differences in treatment of large and small business went still further in the earlier days. In the 1870's and 1880's, large corporations and even more comprehensive business organizations were developing rapidly and exerting their power generally in economic affairs. It was in the transportation field that big business made its most spectacular impact through bringing pressure to bear on rates.

One of the most startling examples appeared in connection with the Standard Oil scandals. In this case it was found that the company was granted refunds below published rates, called rebates, which its competitors were not granted. Further the railroads had to pay to Standard specified amounts for each barrel of oil carried for competing oil companies. This represented the extreme in discrimination between shippers. Discrimination of lesser sorts was not unusual and was a part of the early rate pattern.

Rate determination: Initiation of a rate change. Thus far rates have been considered as being impersonally set by an interplay of inanimate forces. Actually rates are made by individuals, who are officials of the firms in the transportation business. If a shipper or the representatives of a locality want a rate changed or a new one created, they discuss the matter with some local representative of the carrier in its traffic department. This representative's consideration of the request is subject to the review of the higher officials of the department. They are likely to consider the amount of tonnage and revenue that may be attracted to the company, the bargaining strength of the shipper, the effect on other shippers of similar or related commodities, the relation to costs, as well as the reactions of other carriers.

Railroad rate associations. Quite early in the history of railroads, regional associations were established to consider common interests of competitors in rate making. The officials of different companies, as members of these associations, attempted to agree on certain limitations on the freedom of the rate-making process. They wanted to avoid the extreme rate-cutting of the 1870's and 1880's by certain carriers which, for short periods, had forced rates far below the immediate expenses of operation. Another objective of the early associations was to prevent undercutting of rates by billing the shipper for less weight of freight than he actually loaded. Impartial weighing by association representatives provided a check on that. Further it was less complicated and probably more economical to have a single agency publish rates for all the carriers in an area, rather than for each to publish its own.

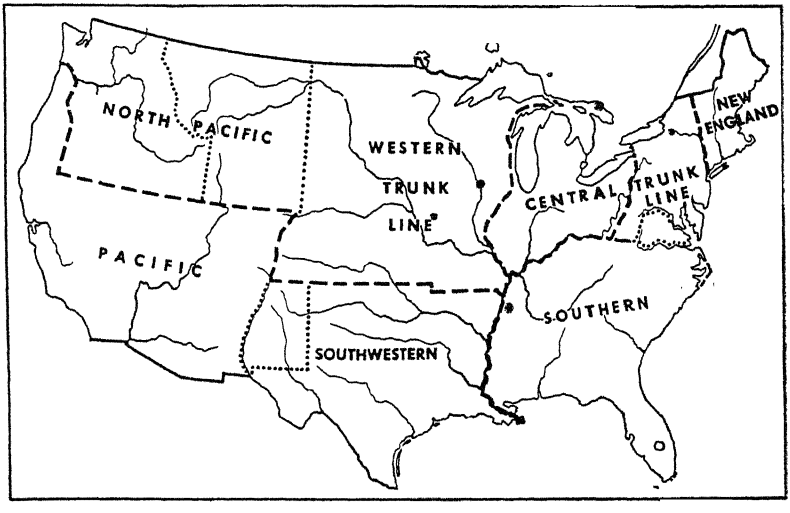


FIG. 43. TERRITORIES OF REGIONAL RAILROAD RATE ASSOCIATIONS IN THE UNITED STATES

The geographic pattern of the associations followed the contours laid out by coverage of the typical railroads in the various parts of the country, the areas being similar to those laid out for the class rate pattern. (Figure 43.)

Because of the early difficulties in enforcing compliance with the limitations imposed by the associations, the terms of association

were made as binding as possible. In particular a member railroad had to agree to abide by a decision as to what rate should be published. After the passage of the Sherman Antitrust Act of 1890, such compulsion on individual carriers was held illegal.

To meet this, the associations changed their requirements, so that their members merely had to agree, before publishing competitive rates, to submit these rates for consideration by the associations. This provision has remained in effect to the present.

The consideration under the modified arrangements may in case of controversial rate proposals involve a lengthy procedure. First a subcommittee may hear the arguments; then an appeal may be taken to a higher committee. From there several further appeals are possible. Finally an individual carrier may go its own way contrary to the association. During the procedure a carrier's immediate competitors can vote several times with respect to its rates and have ample opportunity to use informal persuasion against its proposed action.

All this creates a general atmosphere of uniformity, of "keeping in line." Independent action—the name applied to the final step of a company in going ahead and publishing rates not favorably voted upon by the association—is likely to be frowned upon. The association rests on the principle that "you must in some manner have a meeting of minds as to what the rates shall be," [1] rather than stressing the idea that each company has the potential right to make the rates it considers fit and profitable.

Clearly this involves a distinct deviation from the process of pricing as it is assumed in the theory of pure competition or even imperfect competition. Instead of carriers being perfectly free to make the price decisions which they feel to be for their individual benefit, the pricing process may be delayed while it is subjected to association procedure. Individual interest is subordinated to that of the group, or even to that of its stronger members. Such subordination may restrain a low-cost or aggressive firm from pricing so as to get the maximum share of business, as well as restricting experimentation or innovation in pricing.

Government regulation of particular rates. The results ratewise of the variations in competitive forces and bargaining power, which have been discussed, led to steps toward government regula-

[1] Interstate Commerce Commission Reports. Washington: Government Printing Office, 1923. Vol. 77, p. 271.

tion of transportation in the 1870's and 1880's. Added to these results was the evidence that the railroads were attempting to reduce over-all competition by pooling agreements, to apportion traffic, by the rate associations, which would rigidly bind individual carriers to any rate agreed on by the group, and by overt combinations of competing carriers.

The managers and others interested in railroads fought against government regulation in the legislative halls of the country. They also engaged in long-drawn-out attempts to emasculate the laws through court proceedings. The overwhelming pressure for regulation came largely from the western states, through farm organizations and crusading representatives of the people of those states.

The cornerstone to an effective regulatory structure was laid in 1887 with the passage by Congress of the "Act to Regulate Commerce" and the creation thereby of the Interstate Commerce Commission.

The interstate commerce act laid down four fundamental rules for control of rates by the commission, though it took over twenty years of litigation and modification of the law before that control could be said to be reasonably effective.

The first rule was taken over from the old common-law principles. These had applied rather ineffectively to common carriers, which were those holding themselves out generally to the public as providing transportation service. This rule was that charges should be reasonable and just, and that *unreasonable* and *unjust* charges were unlawful.

The second rule made it unlawful for one person to be charged differently from another for a like service. The wrong committed in violating this rule was designated *unjust discrimination.*

The third rule declared that it was unlawful for a carrier to give *undue preference* to any person or locality or type of traffic or, in reverse words, to subject any of them to *undue prejudice.*

The fourth rule was the so-called *long-and-short-haul clause.* It required that the regulatory body had to approve the cases where a railroad wished to charge more for the haul over a short distance than it did over a longer distance which included the shorter.

These four rules became the bases for the regulation of individual rates for the railroads. With the addition that companies must adhere to published rates, these rules remain the core of rate control today. Furthermore with the exception of the fourth, they

have been carried over into the regulation of other types of carriers which have come under government control.

The purpose of the rules was to limit the irregular contours in the rate pattern arising out of the unevenness of competitive forces. They were to protect weaker shippers or noncompetitive localities. In practice these rules resulted in the elimination of rebates and of deviations from published rates. They helped put the small one-carrier town in a better position with respect to rates.

But there still are important deviations from equality of rates for equal mileage and other conditions. This is because the regulatory bodies have allowed the carriers with different geographic patterns to make rates profitable for their particular coverage of the country. It is also because competition between different types of carriers and different localities has been permitted to continue its influence.

General level of rates: Government regulation: Before 1920. While the original approach of government regulation was directed at particular rate relationships, review of railroad-proposed over-all rate changes inevitably forced consideration of total revenue and profits.

The first instance of this aspect of regulation occurred in 1911, when the railroads proposed general increases. They were denied by the Commission on the grounds that current earnings were higher than ever before, and that the necessity of the increases had not really been demonstrated. Later, in 1914 and 1915, some small increases were approved, and in 1917 a general 15 per cent increase was allowed.

From then till 1920 the general responsibility for rates fell upon the United States Railroad Administration, which had taken over the railroads during the war. In this interim period increases averaged 28 per cent and 18 per cent, respectively, on average freight and passenger charges. Basic passenger fares reached a level of 3 cents per mile.

The 1920 rule. In 1920 the legislation dealing particularly with the regulation of these over-all levels was passed by Congress. In this area a complete break was made from the old approach of control aimed at restraining discrimination and unreasonableness. The regulatory body was given a responsibility to see that rates were so set that the railroads, as a whole or by groups, would earn a fair return. This was hailed as a great advance in rate making.

Ancillary provisions were that the issuance of securities should be

subject to check, and that consolidations of railroads should be fostered by the Interstate Commerce Commission. The latter notion was carried even to the point of suggesting that strong roads take over small weaker ones. New lines, which might upset earning power, were not to be built without authority from the Commission.

Most drastic was the provision that a proportion of earnings above a fixed amount could be taken away by the government. This procedure was technically called "recapture." The funds so collected were to be disbursed for the benefit of weak roads.

Under the new rule immediate increases, varying from 25 to 40 per cent, were granted in freight rates in response to peak wages and prices. The highest increase was for the East, which obtained 40 per cent, and the lowest for the South—25 per cent. Basic passenger fares for the whole country were increased 20 per cent, to 3.6 cents a mile. With the wage and price declines of 1922, the trend as to freight rates was reversed, the Commission on its own initiative lowering rates to the extent of 10 per cent.

The next over-all rate-level consideration did not come for nine years. In 1931 the railroads asked for a general increase of 15 per cent in rates, to make up for the revenue decline of the depression. In response to this, the Commission allowed small increases equivalent on the average to 2½ per cent. With the prices of all other items in the economy falling greatly, even these increases seemed questionable to many, including some railroad officers. There was a real possibility of the railroads pricing themselves out of the transportation market, by moving their prices upward when other prices were being drastically lowered.

By 1933 it became clear that no regulatory body could, by control of the general rate level, insure that total revenue could be kept up sufficiently in all phases of the cycle. Doubts had developed earlier in connection with passenger fares, which had been allowed to remain at the basic level of 3.6 cents a mile since 1920. Even in good times such a fare could not guarantee that the railroads could hold traffic against the alternatives of the private automobile and the bus.

As to insuring a fair return, it became evident not only that revenue could not be maintained, but also that costs were beyond effective control by the Commission. Labor prices had risen during the 1920's, and in the 1930's the possibilities of reductions in im-

portant classifications of workers proved limited. Responsibility for managerial efficiency and service standards, both important elements of cost, could not be assumed effectively by a regulatory body.

With respect to recapturing excess earnings, even in good times, it developed that the possibility of long-drawn-out litigation and the ease of increasing legitimate expenses made recapture impractical.

All in all, thirteen years of experience with the 1920 rule of rate-making demonstrated that it was thoroughly unworkable.

The 1933 rule. In 1933 a new rule of general rate-making was incorporated in the law, which was also to set the standard for the regulatory provisions for motor carriers in 1938 and air carriers in 1940. The regulatory bodies were required to consider rates in relation to three things. One was the effect of rates on the movement of traffic, another was the relationship of rates to the need of furnishing adequate service at the lowest cost, still another was the effect of rates on the revenues needed to provide such service. These were far less rigid than the earlier requirements; in fact they were merely guideposts. But in spite of this, the hazards and difficulties of general rate-level regulation have not been lessened.

Following the change in the law for the railroads, further general increases were requested. In 1935, 4 per cent more was granted on an emergency basis for a short period. In 1938, a 10 per cent general raise was authorized, with but half that amount allowed on agricultural and forest products, and still less on coal. The second great war period saw an initial increase, with no further changes until the postwar reconversion. Then the loss of wartime revenues, combined with increasing labor and material costs, badly unbalanced the financial structure of the railroads.

Historical review of general rate levels. The history of the general railroad rate-level from 1889 to the present provides an interesting summary of the results of the interplay of all factors—labor and material prices, improving techniques, competition, and regulation. The average gross revenues per ton-mile and per passenger-mile are the available indexes of the general levels.

Freight revenue per ton-mile, after a rather sharp decline in the 1890's, varied narrowly around an average of 75 cents from 1898 to 1917. The First World War brought a rise in wages and other prices that was followed by an increase in revenue bringing the average up to a little over $1.25. From this peak there was a gradual reduc-

tion each year through the good times of the 1920's into the decline of the early 1930's, until the level of $1.05 was reached in 1932.

From 1932 to 1934 there was a more marked drop to 99 cents. The decline was halted temporarily by rate increases in 1935 and 1936, but 1937 saw a further decline to 95 cents. The 1938 general rate increases pushed the average back to the 1934 level, but by 1940 individual reductions brought it to 95 cents again; around this level it varied only slightly down to 1945.

Passenger revenues per mile were subject to less marked changes over their history. The averages include commuter revenue at reduced rates, as well as that from all other passenger service. The 2-cent average level was adhered to closely from 1890 to 1918. The First World War saw a rise to 3 cents. A slow decline followed, during the 1920's, to 2.8 cents in 1929. A rapid drop from 1930 to 1936 brought it below the early level to 1.8 cents. Since then the average has varied narrowly around 1.8 cents.

The various forms of transporation at present. The basic elements in the economics of transportation have been discussed primarily as they appear in the field of railroad transportation, because here a long history provides the most available material and shows most factors at work. The modern transportation system, however, comprises far more than railroads alone. While they were operating over 240,000–odd miles of line in 1946, scheduled air carriers were flying over routes totaling 57,000 miles. Intercity bus lines were providing service over more than 1 million miles of the highway system. The highway system in turn was more than ten times the length of the railroads. Its main function was to provide, not for public carriers, but rather for the private traffic of the automobile or truck not operated for hire. In addition there was a network of pipe lines extending over more than 100,000 miles, besides inland waterways, over nine feet in depth, extending over about 10,000 miles.

The capital invested in these facilities provides another index of the composition of the transportation system The figures just before the Second World War were in the neighborhood of 20 billion dollars for the railroads. Valuation of highways is most difficult because suitable accounting for capital expenditures is not carried out. It is probable that the value is slightly less than that of the railroads. To this however must be added highway vehicles to the amount of approximately 11 billion dollars. Waterway facilities

would account for some 5 billions more, and air and pipe-line transportation for at least 500 million dollars each.

As regards passenger travel in peacetime, roughly eight-tenths is by private automobile, something over one-tenth by rail, with the remainder distributed among the other types. The airway portion is rising rapidly, but it is still less than one-tenth of that carried by the railroads.

Freight service as measured in ton-miles is divided roughly into two-thirds rail, one-fifth water, and the rest pipe line and highway. The water portion reaches that magnitude because of the large traffic in coal and iron ore on the Great Lakes and the coastal traffic in coal and petroleum products.

In terms of dollars annually spent for these services in a peacetime year such as 1940, highway transportation accounted for the lion's share, probably of the magnitude of 10 billion dollars. The railroads took 4.4 billions; the remaining carriers added up roughly to another 2 billions. These figures may be compared with a national income at that time of just under 80 billion dollars. The creation of place utility comprises an important segment of the total economy.

CHAPTER XXXVII

Government Regulation of Industry and Trade

Introduction. In a system of free enterprise the results of the activities of those engaged in production come, as we have seen, not from central planning, but in an apparently accidental fashion from a multitude of independent and seemingly uncoördinated decisions of entrepreneurs motivated by the desire for profits. Actually these decisions are coördinated by market prices which guide individual businessmen in their decisions. Indeed private enterprise rests on the principle that market prices, influenced by the demands of consumers for various goods, by the relative costs of producing goods, and by competition of business enterprises, provide a workable automatic regulation of economic activity.

Although competition and competitive price are thus relied upon generally to ensure efficiency in production and prices in equilibrium with costs, it has long been recognized that in some businesses competition is self-destructive and leads naturally to monopoly, and that in certain other situations business activity, restrained only by competitive price, produces results injurious to the interests of society. There are differences of opinion as to the appropriateness of government regulation in certain cases and as to the nature of the controls that should be applied, but the necessity of government regulation in some parts of the economy, as an adjunct to the automatic control of competitive price, has been generally accepted.

Types of government regulation. Government regulation of business has been imposed for various reasons, and with correspondingly various aims. Excluding agricultural, railroad, and public utility regulation, we find that American business regulation to date may be divided into four main categories.

The first main reason for government interference in business has been the belief that the vigor of competition has been reduced and the character of competition modified by certain changes in the economy. The principal changes referred to are structural in character.

One change involves the rise of very large-scale business enterprises, through simple growth or through combinations or mergers, with the result that many industries are dominated by a few large sellers. Such concentration of market control may not only tend to limit economic opportunity for most individuals, but may also result in modifying the character of business competition and the functioning of the price system.

An associated sort of development which may affect competition involves the growth of certain types of behavior among businessmen, both in highly concentrated and in less concentrated industries. One such procedure is the adoption of various techniques or conventions to limit or control price competition. Thus collusive agreements to fix prices and share markets have become increasingly common since the close of the Civil War, and there has been a similar development of less formal means of controlling price, including price leadership by the principal firm in an industry and mutual filing of price information in trade associations.

A complementary development is the tendency to replace emphasis on competitive price by emphasis on competition in the form of product development, advertising, and other sales promotion. Such developments have an effect similar to that of concentration of market control, in that they modify the character of competition and the sort of automatic regulation which competition imposes on the economy. Modifications of this sort have been popularly suspected or disliked, and characterized as being generally "monopolistic" in character.

These tendencies within the American economy have given rise to the first main type of regulation, which attempts to facilitate, reinvigorate, or enforce competition. Such regulation has in general approached this end, either by attempting to limit the degree of concentration in industries, or by declaring illegal those agreements and similar devices which business firms adopt in order to limit competition. The policy of the United States Government in this regard is embodied principally in the Sherman act, the basic antitrust law passed by Congress in 1890. This act has been supplemented in a degree by provisions of the Clayton act and the Federal Trade Commission Act, both passed in 1914. Together the antitrust laws are designed to place some controls on concentration and to outlaw collusive agreements which are "in restraint of

trade." Antitrust regulation has traditionally been the major aspect of the regulation of business in the United States.

A second ground for governmental interference in business has been the desire to eliminate or control certain by-products of the competitive process. The concern here has been not with the lack of competition, but with the fact that competition in certain instances involves questionable practices which tend to be injurious to consumers or to the legitimate interests of other competitors. These difficulties have given rise to a second sort of regulation, concerned with establishing a satisfactory "plane" of competition, through protecting competitors from certain tactics and consumers from misrepresentation and from injurious products.

A third ground for governmental regulation is more or less an antithesis of the first one mentioned. That is, regulation has been resorted to in order to restrain or control business competition which has become "too vigorous."

Legal restraint of competition has not ordinarily been applied directly in the interest of consumers or of the general public, but rather to protect business firms from losses or insufficient profits to which allegedly destructive competition gives rise. It is argued, however—at least by the parties seeking relief from competition—that such restraints are necessary in the public interest, either to preserve a long-run supply of the commodities involved, or to prevent the monopolization of the industry.

A fourth ground for governmental interference in business operations has been certain tendencies in the financing of large enterprises and in the marketing of corporate securities, thought to be operating to the disadvantage of investors in corporate stocks and bonds. These tendencies appeared to be acute in the great depression from 1929 to 1933 and gave rise to the passage of the securities acts of 1933 and 1934 and a public utility holding company act in 1935.

The securities acts regulate the conditions under which new securities may be offered by corporations and also provide regulations for the operations of the stock exchanges through which such securities are ordinarily traded.

The public utility holding company act limits the conditions under which electric and gas public utility firms may be combined through the device of the holding company. It was designed to discourage such financial amalgamations where there was no sound

technological basis for combination, and where financial manipulations might therefore be suspected.

If we ignore the multitude of wartime controls, the four sorts of regulation here discussed encompass the majority of American legislative and administrative efforts to control the operation of business enterprise. There are other issues, however, which have given rise to proposals for additional regulation, and these also must be taken into account. Thus the separation of ownership from control in large corporations and the alleged restrictive effects of the patent law on the freedom of competition are frequently cited as problems.

Anti-trust laws: The problem. The main facet of American public policy toward industry is embodied in the antitrust laws, which are aimed in general at maintaining a desirable degree of competition by preventing (1) "monopolization" and (2) "restraint of trade." Since the basic laws under which this policy is carried out were passed in 1890 and in 1914, it is a policy of long standing and with an extensive history.

The problem of competition is not necessarily unchanging through time, nor was it conceived, in 1900 for example, in the same terms as those in which we now understand it. The movement toward passage of antitrust legislation received its stimulus in the growth of large industrial enterprises and railroads after the Civil War and in the resulting popular agrarian suspicion of the "monopolists" and financiers of the eastern seaboard. The principal manifestations of this concentration are found first in the formation of *pools* and secondly of *trusts* and *holding companies*.

A pool is an agreement of independent producers, whereby their freedom of action in respect to certain matters is surrendered to a representative committee or to the group acting as a whole. A trust, technically defined, consists of the creation of a board of trustees to which are transferred the shares of stock in the several companies, together with the assignment of voting power. In return, the former holders of the stock receive trust certificates, which give the holders a proportionate share of the joint net earnings of all the stock held in trust. Since the trustees hold a controlling number of shares of stock in each company, they are able to elect the several boards of directors and thus attain uniformity of policy and action.[1] The

[1] It will be noted that the term "trust" is used in two senses. From a technical viewpoint a trust is the form of combination here described, but the use of the

holding company, which involves outright ownership by one corporation of stock in one or more other corporations, was the common legal device employed in combining several competing enterprises into a single one.

Between 1880 and 1905, combinations were formed at such a rapid rate and affected so much of American industry as to constitute a real "merger movement," which aroused popular apprehension lest there was coming a degree of concentration approaching monopoly that would result in excessive prices to buyers and excessive profits to sellers. It was also feared that small competitors would be driven to the wall by the power of the large combines. Experience during the thirty years following 1880 provided considerable grounds for these fears. Thus the early conception of the problem of effective competition was phrased largely in terms of the dangers of monopolization to consumers and to small businessmen. The problem was correspondingly conceived as one of preventing trusts and other agreements restraining competition.

In later years it has been recognized that concentration is now a historical fact and one unlikely to be seriously disturbed by feasible regulatory measures. Hence attention has been shifted largely from the issue of whether or not business should be concentrated to an examination of how competition operates in various sorts of concentrated industries.

It is recognized that, where concentration has proceeded to the point of only a few sellers, there may be the possibility of charging monopoly prices and reduction of the opportunities for small business. It is also apparent that the restriction of competition which gives rise to these results may be implemented by collusive agreements or similar conventions among sellers, although whether the tendency to collusion is of such a character that it can really be checked by legal means is doubted by many observers. In addition, it is recognized that the increasing importance of competition by other means than price changes, and the phenomenon of price inflexibility, may constitute serious problems.

The sellers in certain industries producing consumers' goods, although they abstain from active price rivalry, compete rather actively by developing or varying their products, advertising them, and increasing expenditures on distribution and sales promotion.

term to describe any monopolistic combination whether in the form of trust or otherwise is too deeply imbedded in our language and thought to be rejected.

This has tended to raise selling costs or, in other words, to increase the proportion of productive resources devoted to selling and distribution.

In general, sellers in concentrated industries have also frequently refrained from changing their prices in response to fluctuations in demand or cost of production. When such price policies become widespread, the price system may function less effectively in adjusting industry to changing economic conditions. Furthermore industrial price inflexibility puts a heavy strain on the agricultural sector of the economy, when agricultural prices, which are ordinarily quite flexible, decline precipitately during periods of business depression.

The old-fashioned simple problem of monopolistic or concentrated industry has thus become a complex problem of price policies and price behavior in concentrated markets. This situation is further complicated by the fact that, among industries where either concentration or collusive restraint of competition or both are problems, there are several distinguishable categories, which show important differences in type of competition and in the price results of competition.

Provisions of the antitrust laws. The problem of regulating, facilitating, or redirecting competition has always been as complex as it is now, and it was implicitly so when the Sherman Act was passed in 1890 and when its supplements were enacted in 1914. But the full complexity was not then generally recognized.

The popular oversimplification of the problem is revealed in the Sherman Act, the substantive provisions of which are: first, that any contracts or agreements "in restraint of trade" are declared illegal, and second, that "monopolizing" or "attempting to monopolize" an industry is illegal. The Clayton Act of 1914 continued in this tradition by prohibiting certain particular devices or forms of activity which might tend to foster monopolization or substantially to lessen competition. The Federal Trade Commission, an administrative body, in addition to enforcing the provisions of the Clayton Act through investigation and through issuing cease and desist orders, was given the additional power to prevent "unfair" methods of competition, including restraints of trade. Thus, the basic Sherman Act and its two supplements were written in simple, broad, and general terms which reflected a very simplified conception of the problem of business competition.

The effects upon the character of competition, which have re-

sulted from a public policy based on these laws, have obviously been influenced by the drafting, content, and judicial interpretation of the basic acts. Relying principally on the Sherman Act, the federal government has been able to take two sorts of action. It can proceed to require the splitting up of a merger, combine, or large concern, on the ground that it is monopolizing its field; or it can proceed to outlaw agreements in restraint of trade among otherwise independent concerns. Both are accomplished through litigations or suits brought in the federal courts by the Department of Justice against business concerns or individuals who are thought to be violating the law.

Additional powers granted by later acts allow procedure against "tying" or restrictive contracts between sellers and buyers, against price discrimination damaging to a competitor, against intercorporate stock acquisitions which would result in monopolization, and against "interlocking directorates" among competing corporations, where these devices can be shown substantially to lessen competition. These additions, however, have not been very significant in practice.

Intrinsic limitations of the antitrust laws. Given the government's power to proceed along these lines, the maximum efficacy of the antitrust policy must depend mainly upon what changes can be produced in the character of competition by prosecutions against agreements in restraint of trade and by dissolving monopolistic combinations. Many economists in recent years have emphasized the intrinsic limitations of such a policy.

On the one hand, the effectiveness of forcing large companies to split up, and thus lessening the degree of concentration in any industry, is limited by the economies of very large-scale production. Even though in a number of industries there could be more firms of smaller average size than at present without impairing efficiency, there still remains a question as to whether the number of firms could be increased sufficiently to produce any substantial effect on competition.

On the side of collusive agreements, it is generally agreed that restraints of trade are common in modern industry. One school of thought holds that such collusive behavior will not be eliminated by antitrust prosecutions and that, even if it were, competition would not be much different in character. Another group maintains that, even within our considerably concentrated economy, a

vigorous and extensive prosecution of agreements in restraint of trade will secure more effective competition. The main issue concerning the usefulness of the antitrust laws thus turns currently on the effects of prosecutions of agreements in restraint of trade.

Whatever the net effect of such prosecutions, it is evident that, under the present laws, the government can proceed only indirectly against unduly high prices or inflexible prices. The law sets no standards of price behavior; it merely prohibits agreements in restraint of trade. The end results which a free enterprise economy produces can thus be influenced only in so far as the market behavior of businessmen is really modified by a persistent legal attack on restraints of trade. The chance of influencing the height and flexibility of prices in various industries would seem, however, to be greater than that of restricting the volume of expenditure on sales promotion.

Judicial interpretation of the antitrust laws and administrative policy. The actual effectiveness of antitrust policy from 1890 to the present has of course been influenced by factors other than the intrinsic limitations of the basic laws. In fact, the history of American antitrust policy reflects the dominant influence of fluctuations in the vigor with which the acts were enforced by the government and of narrow and changing interpretations of the antitrust laws by the Supreme Court.

From 1890 to 1905, when most of the principal mergers in American industry were being formed, the Sherman Act was not vigorously enforced, and Supreme Court interpretations during most of the period made it appear that even a merger which resulted in substantial monopolization of an industry was not contrary to law.

When the opinion of the court made it clear that such mergers might violate the monopolization provisions of the Sherman Act, a "trust-busting" program was begun. It was pressed with only moderate vigor, however, and was hampered by the fact that most mergers were already accomplished and were difficult to dissolve. Although the courts ordered the splitting up of a few large combines, including the old American Tobacco Company and the Standard Oil Company in 1911 and 1912, widespread action was not undertaken by the government.

Trust-busting came to a halt altogether after 1921, when the Supreme Court modified its interpretation of the Sherman Act and

legalized large combinations so long as they were currently allowing other firms an "freedom to compete" with them.

Thus at first, before 1905, judicial interpretations of the Sherman Act as applied to mergers offered no serious deterrent to concentration and later, after 1920, encouraged a policy of "live and let live" in concentrated industry.

The effectiveness of antitrust law proceedings against agreements in restraint of trade was less influenced by judicial interpretations than by fluctuations in the enthusiasm with which the office of Attorney General enforced the Sherman Act.

Since before 1900 the Supreme Court has been fairly consistent in the stand that any agreement in restraint of trade—fixing price or output, sharing the market, and so on—is illegal, regardless of the demonstrated reasonableness of its results. The principal exceptions to this are found in decisions defining the extent to which competitive firms may go in exchanging price, cost, and market information through trade associations.

During most of the history of the Sherman Act, however, the Department of Justice did not proceed on a very extensive scale against business agreements in restraint of trade. Though such restraints were presumably numerous, relatively few prosecutions were undertaken, and the staff of the antitrust division of the Department of Justice was very small. The force of the litigations which were instituted was lessened, moreover, by the fact that the suits were often undertaken in a conciliatory, rather than a punitive, manner. It is not surprising, therefore, that the restraint of trade provisions of the law have had slight effect on business behavior during most of the period since 1890.

A much more vigorous policy of antitrust prosecutions against agreements in restraint of trade was instituted in 1937. The philosophy behind this new program is that the Sherman Act, long relatively disused and misused, can have a real effect in reinvigorating competition. To get results, prosecutions need to be much more numerous and more vigorously pressed than under previous practice.

Federal Trade Commission activities. The Clayton Act names certain specific activities as unlawful, including tying contracts with customers (designed to foster exclusive dealing with one seller), price discrimination injurious to a competitor of the discriminating seller, certain sorts of interlocking directorates among competing

companies, and the acquistion of corporation stocks by other corporations if the result would be to create an unlawful monopoly.

These provisions, which were intended to supplement the Sherman Act, have in practice been unimportant, for several reasons. First, the act as drafted so hedged the definition of unlawful acts in each category that the scope of the act was limited in practice. In the second place, the court, noting the extremely cautious phrasing of the act, construed it narrowly, practically exempting a large portion of the activities at which the original proponents of the act had purported to aim. Finally, the provisions of the act were to be enforced primarily by the Federal Trade Commission, an appointive administrative agency, which had the power to investigate alleged violations of the Clayton Act, to hold hearings thereon, and to issue cease and desist orders if violations were found. Such orders could be appealed to the courts, and no penalties for violation were involved. This rather mild sort of enforcement procedure did not serve very seriously to deter experimental violations of the act.

The Federal Trade Commission was also empowered, by the act creating it, to proceed via the cease and desist route against "unfair methods of competition." Various agreements in restraint of trade, however, have also been considered unfair methods of competition within the meaning of the act, so that the Commission has supplemented the work of the Department of Justice in proceeding against restraint of trade.

This supplement has, to date, generally tended to be of secondary importance, in part because the Commission could assess no penalties against "unfair methods," whereas Sherman Act litigations could involve fines and imprisonment, together with damages to private parties injured by illegal restraints. The hand of the Commission in such actions was strengthened, however, by the Wheeler-Lea Act of 1938, which established fines for violation of cease and desist orders of the Commission prior to their affirmation by the courts. There is no fundamental reason why Commission action against restraints of trade cannot eventually become an important supplement to the basic antitrust policy embodied in the Sherman Act.

Collusive activity by export associations. Mention should be made of the special status under the antitrust laws of associations of, or agreements among, American businessmen organized for the purpose of carrying on foreign export trade. The Webb-Pomerene Act

of 1918 provides that such associations or agreements shall not be subject to the prohibitions of the Sherman Act, provided that they do not operate to restrain trade within the United States, or to restrain the export trade of any domestic competitor of such an association. Supervision of export associations is vested in the Federal Trade Commission. The purpose of this act was to make it easier and cheaper for American firms to sell and compete in foreign markets. Only a moderate number of industries have taken advantage of the provisions of the act. Where they have done so, however, it appears that their activities have not been closely supervised and that they have occasionally contributed to domestic restraint of trade or to international restraint of trade through world-wide cartels.

The plane of competition: The problem. Although the preservation of active competition has been the main aim of American policy toward business, attention has been given also, ever since the establishment of the Federal Trade Commission in 1914, to regulating the plane upon which competition may be conducted. It is not unknown for competition to take the forms of "stealing" the designs or trade secrets of a rival, bribing his employees, or imitating his brand labels. In the same vein, there may be misrepresentation to buyers of the content or quality of a product or false representation of the dietetic or therapeutic qualities of foods and drugs. In short, fraud, deception, and bad faith may emerge as by-products of the competitive struggle.

Practices of the sort referred to have come to be regarded by most business concerns and by the general public as injurious to the legitimate interests of businessmen and the welfare of the public. Such practices are far from universal in modern business, since a great many industries have set and observed their own standards of behavior. But there is here a field for government regulation which is not necessarily inconsistent with the policy of preserving vigorous and effective competition.

Pure food and drug acts. So far as protection of consumers is concerned, the principal need is in the fields of drugs and foods. The basic regulation here is found in the pure food and drug acts or, specifically, in the Food and Drug Act of 1906 and the Food and Drug Act of 1938. The first law established and made mandatory certain accepted standards for all drugs and also provided for regulation of the quality of canned foods. Adulteration and misbrand-

ing of such products were in effect prohibited. The law of 1938 extended such regulation to practically all foods and also provided standards for the descriptive labeling of foods and drugs. Several acts affecting specific products or industries are also on the statute books.

"Unfair methods" of competition. All other regulation of unfair competitive tactics has been vested in the Federal Trade Commission. By proceeding through investigations, hearings, and orders against various practices as "unfair methods of competition," the Commission has, over a period of time, established what amounts to a set of rules governing the conduct of competition. In so doing, it has essentially extended the common law, which frowned on certain practices by allowing persons injured thereby to recover from the perpetrators.

The records reveal a long series of actions in which the Federal Trade Commission has sought to prohibit misrepresentation, misbranding, and the like. In general, it has been successful in establishing, through judicial decisions, certain rather definite limits on competitive methods. Prior to 1938 the Commission was hampered by the judicial requirement that any unfair practice must be both injurious to a competitor of the guilty party and contrary to the public interest. Also the fact that no positive penalties were provided for unfair practices meant that the law lacked "teeth."

The Wheeler-Lea Act of 1938 greatly strengthened the hand of the Commission in such proceedings. By proscribing, as such, "unfair or deceptive acts or practices" in commerce, it empowered the commission to proceed solely in the public interest against questionable competitive tactics. The Commission was also empowered to seek injunctions against false advertising of foods, drugs, cosmetics, and curative devices, prior to final determination of its legality. Further the act provided fines for violations of final commission orders and made it a criminal offense to disseminate false advertisements of the kinds of products just mentioned, where injury to health or intent to defraud might be involved. With these powers, the Commission has become an active policeman of the conduct of trade.

Restraint of competition: Origins. Although it has been the expressed general policy of the American government to facilitate or enforce competition in the business world, this policy has not been followed without exceptions. In effect there has in the last fifteen years been developing a distinct, though as yet limited, counter-

line of policy in the direction of restraining competition in certain industries or in certain circumstances.

Because such measures run counter to the widely held belief in the desirability of competition, the intent of these regulations is generally somewhat disguised. An attempt is sometimes made to identify them with the general antimonopoly policy, by labeling them as attempts to preserve "fair" competition and thus to check the ultimate overconcentration of industry in the areas affected But we may truthfully designate these regulations as essentially anticompetitive in character.

The origin of anticompetitive business regulation has been found in the demands of businessmen. This is not strange, for profit-seeking enterprises within the capitalist system have never consistently welcomed free, open, and vigorous competition. Competition tends to reduce or eliminate profits, especially in times of stable or declining purchasing power. There is a natural tendency to seek limitation of competition sufficiently to protect profits. This tendency is the basic source of collusive activity in restraint of trade, and it has served as one of the motives for forming large combinations of competing firms.

The belief in competition is still clearly ascendant in American public opinion, and concerted and effective demands for legal restraint of competition have therefore become vociferous and effective only at particular times or in particular industries. Only once in the history of American business did a general demand for restraint of competition become effective. This was in the depth of the great depression in 1932 and 1933, when a prolonged shortage of purchasing power had stimulated violent competition in a great many areas where such intensive rivalry was uncommon. The result was the National Industrial Recovery Administration, a comprehensive plan for limiting competition. Since the demise of the NRA in 1935, restraint of competition along similar lines has been sought and obtained in two fields—bituminous coal and the distributive trades—where especially acute competitive problems existed.

National Industrial Recovery Act. The depression beginning in 1929 was extended and severe enough to impose losses generally on nearly all business and to effect a serious deterioration in wage and hour conditions for labor. Conventional government measures to initiate a recovery had been quite unavailing, and more desperate remedies were sought. In this setting organizations of large-

scale business diagnosed one of the difficulties as excessive and destructive competition and suggested temporary suspension of antitrust laws and adoption of temporary agreements to restrain competition and restore profits. Similarly labor groups recommended legislation to maintain wages and shorten hours, in order to relieve labor from the pressure of depression conditions.

These proposals gained fairly wide support, which was augmented by a spreading popular sentiment, rather vague in its formulation, that the prolonged depression stemmed from basic shortcomings of the competitive system.

In response to these sentiments, one of the first "recovery" laws passed by the new Congress in 1933 established the National Industrial Recovery Administration. The principal provisions of the basic law, as expressed and administered, were twofold. First, every industry was required to draw up and submit to NRA a "code of fair competition," which, when approved, would become law for the industry. Second, such codes were required to include provisions establishing minimum wages and maximum hours for labor in each industry. The real content of the new law thus emerged from the character of the private codes which were submitted to and approved by the government. In essence the country adopted, for the period of a depression emergency, a system of regulating competition through governmentally supervised private industry agreements, with certain counterbalancing guarantees to labor.

A brief characterization of the whole NRA experience is difficult, because of the great variety of individual code provisions and because of variations in experience with administering and enforcing them. In general however the following may be said about the content of the codes, about administrative experience, and about the effects of the law:

Labor was eventually protected largely through a standard wage and hour formula, which was applied, subject to certain adjustments, to the codes of all industries. The fair competition provisions of the codes were generally designed to control or discourage price competition, to control output, and on occasion to share the market among sellers. Such ends were in some cases accomplished by direct agreements on price or output, but more frequently they were approached indirectly through a maze of provisions controlling competitive tactics.

Administrative experience varied among types of industry. Con-

centrated industries with a history of coöperation among sellers generally obtained good compliance with codes, but atomistic and traditionally competitive industries were plagued with difficulties of administration and enforcement.

The effects of NRA on the economy were generally to restrain competition, to limit output relative to any given monetary purchasing power, and to raise prices and profits. It also served to alleviate substandard labor conditions.

The NRA program probably did not contribute significantly to recovery, and it may very well have retarded it. It is interesting to note that the experiment in question was essentially one in over-all limitation of competition by private business, subject to government supervision, and that "government planning" played an insignificant role in the whole scheme.

The NRA was brought to an end in 1935 by a Supreme Court decision which found the basic law unconstitutional.

Competition in the soft coal industry. Experience of that interval when the law was in effect, however, influenced a number of industries in their subsequent thinking on competitive problems. One of these was bituminous coal, where there was a competitive problem of rather long standing and of somewhat special origin.

The soft coal industry had faced a steadily declining demand for its product since 1920. This had exerted continued downward pressure on mine operators' earnings. Ordinarily this would have resulted in eliminating many firms from the industry and eventually in balancing demand and supply at an equilibrium price. As a matter of fact, firms had not left the industry, and there was a persistent and growing pressure of excess capacity in coal mines. This was partly because of the very long life of "fixed plant," once a mine was opened, and partly because of the indisposition of coal-mine labor to leave the mining towns under the pressure of declining wages.

Operating with this chronic excess capacity were several thousand small mining firms, among which there was no possibility of voluntary restraint of competition. The pressure of free, active, and violent competition thus continually drove prices to unprofitable levels, and this pressure on price was reflected in competitive wage cutting by mineowners.

The Guffey-Vinson Act to regulate the coal industry. This sort of competition was therefore held by both business and labor groups

to be "destructive" in character, and indeed it appears that in this particular setting competition was unfavorable to the general public interest. After an unsuccessful effort to maintain coal prices and wages through an NRA code (which of course lapsed in 1935), permanent legislation was sought to regulate competition in the industry.

This finally took the form of the Guffey-Vinson Act of 1937 (an earlier version from 1935 having been found unconstitutional). The act generally provides for a Federal Coal Commission, which is empowered to fix minimum coal prices for all areas and mines, on the basis of average costs of production as submitted by mine-owners. Such costs are adjustable for any wage changes negotiated between mine-owners and labor. Rights of collective bargaining were guaranteed to coal-mine labor, although wages were not fixed by the act.

The effect of the Guffey-Vinson Act is evidently to eliminate competition and to replace it with government price administration, guaranteeing a "fair" return or profit to firms producing soft coal. This is a specific legal protection to business in this field. By inference it also protects coal-mine labor very considerably, since it assures coal-mine owners that, practically regardless of what wages they grant to labor in the field, they will be able to get prices high enough to cover these wages and yield a fair profit. The act therefore goes considerably further than merely discouraging competitive wage cutting.

Inherent in the act is the tendency to protect possible inefficiency from the penalties which a free competitive market would assess. It also tends to encourage labor to force exorbitant wage demands, knowing that any wage obtained will be made economically tenable to the mine-owners through government control of coal prices.

Competition in the distributive trades. Rise of mass distribution. In the distributive trades there was at work a process of structural change which was resulting in the gradual elimination of certain types of firms which had traditionally held a place in the industry. The industry as a whole was not suffering chronic losses, but a certain category of firms was being squeezed out.

The structural change involved was in effect the rise of mass distribution—wholesaling and retailing through chain-store organizations. This movement gained considerable momentum during the 1920's, in the grocery, drugstore, auto accessory, and other lines, and

was proceeding very rapidly in the early 1930's. In general the movement involved the rise of large-scale organization in the marketing field, and it paralleled a similar development in manufacturing industries some forty years earlier. Both horizontal combination of many similar units into a chain, and vertical integration of wholesaling and retailing functions into a single firm were occurring.

In moving into the distributive field, mass distributors did several things which were damaging to the interests of old-line wholesalers and retailers. They frequently succeeded, through large-scale buying power, in purchasing goods from manufacturers at lower prices than smaller buyers could obtain. They undertook to perform their own brokerage and wholesale functions, thus dispensing with the services of old-line brokers and wholesalers, at the same time further reducing the costs of goods on the shelves of their retail stores. Further they changed and developed the technique of merchandising, by streamlining the retail outlet, modernizing its management, and eliminating certain relatively expensive auxiliary services, like charge accounts and free delivery. The economies thus obtained served as a basis for a generally low-price policy by chain stores, involving prices often low enough to cut sharply into the sales volumes of the older "independent" retail stores, and correspondingly to impair the welfare of the "independent" wholesalers.

The immediate plight of such independent distributors was made more serious by the depression which began in 1929, and their resentment of the chains was thus intensified. In the NRA codes for the distributive trades, organized independent interests sought, and in considerable measure won, code provisions which restricted the chains from competing so effectively with them.

With the end of NRA, these organized interests, through their associations and lobbies, turned attention to securing permanent legislation to restrain competition in the distributive trades. Such legislation was sought in the name either of fostering fair competition or of preventing the development of monopoly. As to the latter, it should be noted that it was to be checked mainly by restraining and weakening competition.

Chain-store taxation. The most direct attack on chain-store competition has been embodied in simple chain-store taxation—punitive levies against multiple-unit marketing organizations. A tax per

store in any chain is assessed, which becomes larger (per store) as the number of stores in the chain becomes greater. Such legislation originated as early as 1927 in three southern states, but it progressed most rapidly in the later 1930's, so that at the outbreak of the recent war twenty-eight states had chain-store taxes. Attempts of the anti-chain bloc to secure similar legislation on the federal level were unavailing. Such taxation of course represents a clear attempt to weaken mass distributors and thus to subsidize small independent wholesalers and retailers.

The Robinson-Patman Act. The other legislation which was obtained to protect independents from chain-store competition is of a less direct but not less effective character. The two principal federal enactments are the Robinson-Patman Act (1936) and the Miller-Tydings Act (1937)—both amendments to earlier antitrust laws.

The Robinson-Patman Act was aimed at reducing the advantage which chain-store organizations have obtained in purchasing goods from manufacturers at very low prices. In effect the act prevents such suppliers (or any seller) from charging different prices to different buyers for the same goods, where the effect of such price differences is to lessen or prevent competition between the purchasers, unless such price differences are based on corresponding differences in cost. (The earlier Clayton Act had not defined such price discrimination as illegal.) A manufacturer of canned goods, for example, cannot sell a given product to competing chain stores and to independents at different prices, except in so far as the cost of supplying them is demonstrably different. The Robinson-Patman Act also controls the payment of certain special allowances and rebates, which had been used to reduce the net price at which mass distributors could buy. Enforcement of the act is placed in the hands of the Federal Trade Commission.

The intent of the act is to enhance the competitive position of independent distributors by denying to chain stores those advantages in cost of supplies which resulted simply from their great bargaining power. The effect of the law has been limited, however, by the ability of mass distributors to monopolize certain sources of supply, taking the entire output of a manufacturer or even acquiring his plant.

Miller-Tydings Act. The Miller-Tydings Act approaches the problem of competition in the distributive trades from a different angle. By itself it is simply an amendment to the Sherman Act,

which declares that resale price-maintenance contracts which are lawful in the states where resales are made shall not be illegal under the Sherman Act or the Federal Trade Commission Act.

A resale price-maintenance contract is a contract between a manufacturer or other supplier who puts his trade-mark on a good and the distributor who purchases this good—a contract which obligates the purchaser to charge a specified price for the good at resale. Such a contract would thus require a drugstore to sell a certain brand of mouthwash at a specified retail price. The ostensible justification at law for such contracts is that they enable a supplier to protect the value of his brand or trade-mark from damage through price cutting.

Such contracts had frequently been voided by the Supreme Court, on the ground that they resulted in suppressing competition among resellers and were thus contrary to provisions of the Sherman Act. There was evident in decisions of 1936 some tendency to reverse this stand if state laws governing the case specifically permitted resale price maintenance. Congress proceeded to clinch the matter by specifically exempting resale price-maintenance contracts from the prohibitions of federal law, provided they found support in state law.

State "fair trade" laws. The Miller-Tydings law was thus in effect simply an action taken at the federal level to facilitate the operation of state laws bearing on the same subject. These laws were at the heart of the anticompetitive movement and were known as the state "fair-trade" laws.

Such laws (the first was passed in California in 1931) legalized resale price-maintenance contracts on substantially any branded or trade-marked commodity. Any retailer was thus enabled to enter into a contract with a supplier to maintain retail prices at a stipulated level. These laws also provided, through the so-called "nonsigner clause," that such contract prices were enforceable upon all sellers of the commodities covered, regardless of whether they had entered into such contracts. Selling at a price below that stipulated in a competitor's contract was "unfair" competition and grounds for a damage suit by such a competitor. In effect therefore retailers desiring resale price maintenance were enabled to enforce retail price fixing and elimination of price competition upon an entire community of distributors.

The questionable legality of the fair trade enactments was cleared

up by favorable court decisions in 1936 and by the Miller-Tydings Act in 1937. Thereafter, most state legislatures, under the pressure of organized druggists and other retailers, rushed to pass "fair trade" laws, until forty-four states now have them in effect.

These "fair trade" laws are a form of pressure group legislation. They represent a distinct break with the traditional American policy of encouraging free and active competition and are representative of the tendencies inherent in pressure-group legislation. Their effect is clearly to stifle price competition and to offer legal protection of price and profits to a single sector of the business community. By the same token, they tend to stimulate competition in features other than price, to attract excessive capacity and man power to the distributive trades, and to pass on the cost of these excesses to the consumer.

It is to be noted that the state "fair trade" acts have not been completely effective. Mass distributors and others can avoid their provisions by selling private brands and by other devices, and they have shown some tendency to do so. The threat of damage suits for "unfair" price competition is not always strong enough to deter all such competition. Nevertheless the acts have had substantial effect.

Appraisal of governmental regulation. From our discussion in the preceding pages, it will be evident that the central theme in the United States of public policy toward business has been the preservation of effective or workable competition. There seems to be slight disposition in the United States to follow the pattern established in several European countries of encouraging businessmen to combine further or to establish a comprehensive system of private cartels to regulate themselves. Foreign experience with such a policy is discouraging, and in any event such a policy is contrary to strong traditional American views on economic matters. Equally against American tradition is governmental participation in the making of business decisions, as well as the nationalization of industries.

The antitrust laws, and especially the Sherman Act, have served as the principal basis for facilitating or enforcing competition. This can be done under these laws either by limiting the degree of concentration in an industry or by attacking and attempting to eliminate agreements and conventions which limit or prevent effective competition. The Sherman Act has proved to be not entirely suited to accomplishing limitation of industrial concentration. As

drafted and as interpreted by the courts, none of the antitrust acts enable the government to pursue a systematic and rational policy in controlling "bigness" in business or in preventing undue or uneconomical concentration.

The antitrust laws can be more effective as applied in prohibition of agreements in restraint of trade. They will have maximum effect, however, only if they are vigorously and continuously enforced over the entire economy, without prejudice, exemption, or special favor.

Accomplishment of this end may require also some modification in the patent laws. Patent grants, which bestow seventeen-year monopolies on new devices, techniques, or processes, have been occasionally used to foster a restriction of competition not contemplated in the basic law.

Working directly counter to the antitrust laws is the essentially anticompetitive legislation affecting the distributive trades, which legalizes collusive determination of price in this area. The antitrust laws could be made dead letters if the tendency to exempt industries from the general rule of free competition were much extended.

Government industry: Unquestioned types. Every government is inevitably engaged in industry to some extent. A people may feel that the postal service is too important to the general welfare to leave in private hands. Military considerations may lead a government to control the railroads, the telephone, the telegraph and wireless service, as is not uncommon in Europe. Protection against disease sometimes requires city governments to take a hand in such enterprises as milk stations, bath houses, housing projects, etc. To protect the country against flood and drought, as well as to provide for future timber requirements, a government may establish national forests and build reservoirs and dams to regulate the flow of water. Enterprises such as the Panama Canal or Boulder Dam may be too vast or the returns too uncertain to interest private capital; government, however, may decide that regardless of the financial return such undertakings are worth-while in the public interest.

These examples—and many more will at once occur to the reader—illustrate the fact that there is a province for government industry which is little disputed. Few question the propriety of government industry undertaken in the realm of its primary functions, nor is there much criticism of government projects in fields

where the risk is so great as to deter private capital from entering, when once it is settled that the project would truly serve the public interest. Serious disagreement regarding government industry arises when government invades fields which tradition has reserved for private capital.

Issues involved: Adequacy of governmental regulation. Those who advocate government industry in order to protect consumers from the exactions of private monopoly are not to be confused with those irreconcilable critics of the present order who are persuaded that the weaknesses of private capitalism are so fundamental that the only hope of stable prosperity lies in the direction of socialism.

On the contrary the former advocates of government industry are men who believe in the present economic system and want to see it perpetuated. They are, however, so impressed with the weaknesses of government regulation that they feel that the system can best be preserved by removing ffom private hands those parts of it which have become monopolistic in character and marked by conspicuous abuses.

Opposed to them is a large group by no means convinced either that government regulation is futile or that the abuses of private business are sufficiently serious to warrant so drastic a step. Moreover many of these are convinced that government industry tends usually to be less efficient than private business, and they are genuinely fearful lest an enlargement of the scope of government industry bring consequences even more unwholesome than those which it is intended to cure.

Savings on interest and profits. Attention is directed frequently to the large sums paid for interest on private capital and the large profits made by private enterprise, and it is claimed that these might as well accrue to the whole people in the form of lower rates. It is evident that such claims cannot be substantiated.

So long as saving is left to private initiative, interest is a charge which must be met whether the industry be public or private. It is a payment essential to the accumulation of capital, and the government cannot get capital without the payment of interest any more than can a private entrepreneur.

If attention be called to the fact that government bonds bear a lower rate of interest than industrial investments, the answer is that the whole advantage would tend to disappear in proportion as

government became involved in the risks of industry. If the taxing power be relied upon to make good the losses of unsuccessful government enterprises, this is merely robbing Peter to pay Paul.

Even in a different social order in which capital were not privately owned, there could be no capital without previous saving by someone. Government might, it is true, get its capital without paying interest, but the ultimate cost of saving would not be thereby avoided.

The same general considerations apply to the alleged saving of profits. Profits, as we know, are paid for the service of risk-bearing, and we have concluded that the payment, considering losses and gains together and taking industry as a whole, is distinctly moderate. If the risks of industry were transferred from the shoulders of private entrepreneurs to the government, they would still have to be paid for. Governments cannot take the profits and avoid the losses. The government industry would have to charge enough to enable it to make experiments, assume risks, and stand the inevitable losses. There is no reason to believe that the cost of risk-bearing would be any less if undertaken by the government, unless it were to avoid risk by permitting industry to stagnate.

While the above considerations apply to industry in general, it is possible that in the special fields into which the municipalities particularly have ventured there may be a saving to the consumer by reason of a reduction in the volume of profits. The purveying of water or electricity is generally a monopoly whether the plant be publicly or privately owned. As compared with competitive manufacturing, for example, the risk is slight and in many cases perhaps nonexistent. The stockholders of a privately owned public utility might not be content with profits only sufficient to compensate for the risk involved.

With private ownership pressure is constant to maintain existing rates or secure higher ones. Quite commonly a rate which at one time yielded reasonable profits has, with a thickening of the population, given very great profits. Under private management the action of the public service commission in reducing rates is slow and halting; under public ownership, assuming for the moment that public management be as efficient as private, or even slightly less efficient, either the consumers would get the benefit at once in the form of lower rates or the taxpayers would benefit from lower taxes.

Corporation abuses. In their advocacy of government industry, some lay greater stress on the necessity of eliminating certain undesirable aspects of corporate control than on the actual savings which government industry might secure. They point out that the problem of watered stock, which frequently affects the rate base and therefore the cost of service to consumers, would be done away with, that the looting of utilities by insiders would be impossible, and that speculation in securities of utilities would no longer prevail.

But more fundamental is their belief that regulation of public utility rates has been halting and ineffective and that it is idle to place reliance upon regulatory commissions to secure rates which are only reasonably remunerative. It is not within our province to discuss the problem of regulation of public utilities, but it is perhaps not out of place to note that in many states appointments to regulatory commissions have been on a political basis, that the task of proving that a given rate is unwarranted involves long-drawn-out and very expensive litigation, and that the problem of regulation has been exceedingly complicated because of the interstate character of many of our holding companies.

It is however possible to recognize the existence of these conditions, as do most authorities, without reaching the conclusion that government industry is the only solution, for the maximum possible effectiveness of regulation has certainly not been reached.

Relative efficiency: Financial results. An investigation into the efficiency of government enterprises gives a picture that is far from clear. In one country state-owned railways may give excellent service and show satisfactory financial results, while in another country poor service and continued losses follow year after year. The United States Post Office, the nation's most conspicuous example of government industry, has operated for years without any real accounting system and with a very large and unnecessary deficit. These results are chargeable not only to legislation by Congress but also to serious lack of sound business judgment in the management of the enterprise.

In the field of municipally-owned public utilities the evidence as to financial results is often incomplete, conflicting, and even unreliable. Frequently the accounting systems of municipal enterprises are not adequate or reliable; the United States Census Bureau complains regularly of the untrustworthy accounts upon which it has

to base its statistics. Apparent net operating revenue must in many cases be reduced because of failure to make proper allowances for interest, depreciation, or amortization. Services, such as street lighting, may be provided free to a city without an appropriate charge on the books, and similarly many services may be provided by other departments of the city without appearing on the accounts.

In the face of these facts any sweeping generalities would be unwarranted and unsound. On the whole the financial record of public industry is unfavorable; there are conspicuous cases of financial success, but the number and magnitude of the industries operated at heavy deficits is more impressive, and in most cases—in America at any rate—the actual facts are far less favorable than the published figures on account of the prevalence of faulty accounting methods.

Political interference. There are certain positive weaknesses which seem inherent in government management. In the first place, there is weakness in personnel and discipline. Politics controls more or less the selection of employees, both managers and subordinates. Politics also interferes with discipline; those in charge hesitate to enforce adequate control over their subordinates or to discharge the incompetent, for fear of political protest. For the same reason the managers have often been prevented from installing approved labor-saving machinery, cost accounting systems, or up-to-date business methods. Political sentiment also tends to prevent the payment of salaries sufficient to attract into the public service the most capable managers and technical experts. Thus is efficiency sacrificed and industrial progress hindered.

On the other hand, certain private utilities are frequently in politics and exerting pressure on legislation through lobbies and organized and "educated" public opinion. Society would gain something by removing this danger to democratic government, although it would probably stand to lose far more from the development of bureaucracy and the political control of government industry.

The motive in business. Public industry is weak in comparison with private industry as to the motive of the managing head. The private businessman puts the utmost of his skill and energy into his business because he has a personal stake in it. Even if he be only the hired manager of a corporate business, he finds the same standard held up to him by the stockholders, who look to him to protect their capital and produce profits for them. His position and

his fortune depend on making a success of the business under his charge. Self-interest is the spur which drives the private businessman to make his business efficient.

To the public official this motive is weak or absent. He has not risked his own capital in the government industry. If profits are made they will not accrue to him, and he will not have to bear the financial consequence of failure. The continuance of his position and his further advancement do not depend upon the judgment of a critical body of stockholders as to the efficiency with which he has managed the business. His position is more often dependent upon pleasing the general body of voters or the leaders of his political party. This approval may be won by easy-going treatment of his subordinates and showy service to the public, rather than by the hard test of business efficiency.

This is of course not to assert that there never is nor can be efficient management in public business. There are capable men animated with a keen sense of public service, and there are communities in which the citizens insist on able and honest management. There have been conspicuous examples of high-grade leadership impelled by motives of patriotism and public service, especially in times of national emergency such as war.

On the other hand, it is not to be denied that there are instances in private business where other considerations complicate or interfere with the simple relationship between efficiency and the profit motive. This is most likely to be found in the very large corporations or combinations where ownership and management are less closely identified, and other interests than those of the owners are more likely to appear. Indeed as private business units become exceedingly large, there is sometimes observable a tendency to take on certain of the characteristics of government industry. Considerations such as these however are the exceptions to the general rule as to the important difference in motive between the managers of private and of public industries.

Conclusion. The problem of public *versus* private industry must still remain undetermined, in the sense that it is impossible to arrive at any single broad generalization as to the relative merits of the two forms of ownership. It is not at all likely that the problem ever will be solved in the form of such a general conclusion. In no two countries and with respect to no two classes of industry are the conditions of the problem the same. The decision as to any

industry may go one way in Germany, the opposite way in America; in any given community the decision may be for municipal water works and against municipal gas works. The defects of private industry may compel public ownership in one case, as of forests on the water sheds, while the weaknesses of public ownership may lead to substitution of private industry, as in the case of the Philadelphia gas works some years ago. In one state effective regulation may give satisfactory rates and service, while in another a weak commission may lead municipalities to assume the risks of government industry.

The mass of evidence and the body of general principles bearing on the question are such as to place the burden of proof upon the one who advocates government ownership. From this position the public will undoubtedly feel its way along the road of economic progress, determining each case upon its merits in the light of such knowledge as may be at hand.

One general conclusion may perhaps be stated with some confidence. The public is not likely to tolerate unlimited private monopoly in the necessities of life upon any large scale. In any basic line of industry where monopoly appears inevitable, the choice is between private industry with government regulation and government industry. Attempts at government regulation have thus far been generally short-sighted and bungling. We may now be upon the threshold of the first real test of this program; should it fail, the alternative of government ownership would still be open. In the meantime there may well be no definite conclusion—no decided swing either to public or private ownership—but rather a balance between the two, with private ownership the fate of the inefficient public plant, and public ownership a threat and a spur to the privately owned public utility.

CHAPTER XXXVIII

Contrasting Economic Systems

Introduction. Present status of the free economy. This book has been devoted to the study of the kind of economic order which exists in the United States of America, one which relies in the main on voluntary gain-seeking activities of individuals and groups to supply the economic needs of the people. In the course of our study, we have seen that the system is not one of unlimited individual freedom, nor one which relies solely on the profit motive to provide necessary goods and services. Competition is restrained at many points by the regulatory powers of the government in the interest of general welfare, and in certain areas of national life governmental bodies have taken over productive functions.

But these regulatory and productive activities of government are exceptions to the prevailing rule, and although they have been on the increase in recent years, they still play a minor role in the economic life of the nation. Bearing in mind these exceptions, it is proper to call this type of social order a system of free enterprise. Another name applied to it, particularly by those who seek to change it, is capitalism.

In the world at large, systems of free enterprise are not the usual, but rather the exceptional, structure of civilized nations. This type of economy is in operation today only in a few countries—the United States, some of the Dominions of the British Empire, a few of the Latin-American countries, one or two small states in Europe. Throughout the rest of the world, embracing a large majority of the nations and the preponderance of human beings, economic life is organized along very different lines. Moreover the area of the world served by the system of free enterprise has been narrowing steadily in recent years, as one country after another has abandoned this type of social order and adopted another.

Every country, including the United States, which is served by a system of free enterprise, contains organized movements among its people working for a complete reorganization of the social struc-

ture. The emergence in other civilized countries of radically different social systems has been due to the success of similar revolutionary movements in gaining control of political power.

Despite diversities in the details of their programs, all these movements—at least all of any importance—derive their ideas and their motive power from the same body of doctrine, the original doctrines of socialism. Hence we shall group them all under the general title, *socialism,* reserving for special consideration an important section, *communism,* which, though originating in this same body of doctrine, has broken off from the other socialist groups. Some attention will also be paid to *fascism* because of its close connection with revolutionary socialism, although it is often and erroneously classified with the conservative movements of our time.

Socialism: Theoretical basis. All socialist doctrines contain four groups of ideas: an indictment, an analysis, a program of reform, a campaign. The *indictment* is a list of charges against "any and all industrial systems based on private property and competition." The *analysis* traces the development of capitalism back to its origin, dissects its present structure, and stresses certain developmental tendencies which, so it is argued, must inevitably destroy our present order. The *program* outlines a substitute for the capitalistic system which socialists believe will remove the evils of modern capitalism. The *campaign* is a plan of immediate tactics for those who adhere to its cause.

There is greater unanimity of opinion regarding the first two branches of doctrine than the last two. In their indictment of capitalism, socialists differ only in the breadth of their condemnation of the present order and the relative stress they place on its different faults. Since the socialist analysis of our institutions is derived almost entirely from Karl Marx's classic interpretation of capitalism, there is considerable agreement here.

Socialist programs differ widely in detail, but we can discern an essential similarity among them. They all propose a type of society in which there will be no private property in wealth used for productive purposes, and they agree further that the productive system of the future shall be governed, not by competition and private initiative, but by some sort of collective control and guidance. Socialism then is destructive of these two basic institutions of our order, private property and free competition in enterprise. The programs differ in their substitutes for these institutions.

In regard to their campaign, socialists divide into two groups: those relying on peaceful, evolutionary methods, and those who place their reliance on violence and force. The former are *constitutionalists,* believing that gradual increase in their power on the political field will eventually bring them to their goal by thoroughly legal means. The latter are *direct actionists;* they have no faith in politics, being convinced that only through force or threat of force will the revolution come.

Any definition which attempted to embrace all these phases of the socialist movement would be too cumbersome for practical use. We shall define the movement in terms of its program, stressing the essential similarities between the different socialist schools. *Socialism is a program of reform which repudiates private ownership of the means of production and competitive control of industry; it proposes a type of society in which productive capital will be owned collectively and economic activity will be controlled by authority.*

Socialist indictment of capitalism. The details of the socialist indictment may be summarized with reference to the institutions of competition and private property. Competition is condemned as a method of supplying the economic needs of society, because it is wasteful of resources and effort, because it is unstable and erratic, because it results in strife within and between nations. The institution of private property is condemned because of the inequality and injustice it produces and because of its seriously adverse effects upon the processes of production. The two institutions are jointly condemned because they destroy human liberty.

The socialist charges against the competitive system on the score of its wastefulness have long been voiced. The point is stressed that individual enterprise results in duplication of effort. Competitive advertising, competitive salesmanship, and other forms of competitive activity are substantial costs of each enterprise. These costs, the socialists affirm, could be reduced if the services common to all producers were provided through a common agency.

In the competitive system the superior business methods and most efficient machines are controlled by individual competitors for private profit. The socialists claim that increased efficiency for society as a whole would result if the best processes were made the common possession of the entire industry. Thus far the argument bears some

resemblance to that which contends for the superior efficiency of combination and large-scale production.

In a more general sense the socialists urge that the present system is wasteful because entrepreneurs in their search for profit direct capital and labor into fields of activity which are already well enough equipped to serve the needs of society. It would be more economical, they say, to use collective intelligence in forecasting our needs and conscious coöperation in adjusting the means of production to the ends in view.

It is urged also that the uncontrolled pursuit of private gain exhausts irreplaceable natural resources and destroys human productive energy, as when it puts children to work prematurely and throws old people into idleness while still able to perform work.

In charging the economic order with instability the socialists develop an explanation of the business cycle which makes competition its basic cause. Competition produces anarchy in the productive system and, when coupled with inequality of income, results in overproduction. Periodically the system collapses through its inability to market its product, and there ensues an incalculable loss from idle capital and labor. These cyclical catastrophes, it is asserted, could be avoided if competition were supplanted by a planned economy, controlled and guided collectively by authority and force of the state.

In charging that the competitive order encourages strife, it is pointed out that employers and workers strive with each other over division of the income from industry. Producer strives with producer for advantage in the market. Out of this strife spring practices which result in loss of productive forces. The anomaly of the situation is that these contesting parties are really coöperating toward common ends. Intentional coöperation is less wasteful than antagonistic coöperation.

From a broader point of view, socialists believe that the competitive system is the root cause of international war, the striving of nation with nation over markets. They believe that their system will automatically produce harmony between individuals, social classes, and nations.

The charges against private property are in part implied in the foregoing attack on competition. But this institution, as the socialists view the matter, has defects of its own; the great inequalities in

incomes are caused by private ownership of income-bearing wealth and perpetuated by the institution of inheritance.

The socialists believe that capacity is distributed with approximate equality among the different income classes. Opportunities for self-development cannot be equalized so long as some men are born to a position of privilege while the many are born in an environment of poverty, ignorance, and hopelessness. This situation, they say, is not only unjust on ethical grounds but is inefficient and precarious; inefficient, because it smothers the latent abilities of those in the lower groups and subsidizes idleness in the upper classes; precarious, because it arouses the envy and resentment of the victims of the injustice, who are an overwhelming majority and have the power to destroy the system.

Inequality is particularly vicious because it distorts the whole productive system out of alignment with the real needs of society. Only those things are produced which will sell at a profit. Hence the kinds and quantities of goods resulting from the combined activity of mankind are determined by the spending power of individuals and not by the test of what is needed to nourish a strong, healthy population. Many things of vital importance to human welfare are not produced at all because no individual entrepreneur can make a profit out of their production.

The socialist system proposes a quite different basis of production. The social group will decide what things are needed to support a wholesome life, and these things will be produced first in such abundance that there is plenty for all. Only surplus labor and capital will be devoted to the production of luxuries and extravagances, while injurious goods will not be produced at all.

As a final and in their opinion most devastating charge, the socialists condemn capitalism as destructive of human liberty. The individual liberty of which so much is made by the defenders of our order is dismissed as a myth. To the socialists the present order means the enslavement of the great majority of propertyless people by the owners of wealth and the subjection of the whole body of society to the dictates of those who dominate its essential industries.

The wages system is condemned as a species of slavery. It is explained as an arrangement between the few who own the instruments of labor and the resources of the earth and the many whose only hope of survival consists in their getting access to these things.

The outcome of the arrangement is that the many deliver themselves into the power of the few, their productive energies to be exploited to the profit of the masters.

The economic destiny of the social group as a whole is, according to this indictment, at the mercy of the more ruthless or more fortunate individuals. The landlord exacts his tribute for the privilege of using the gifts of nature. Those who have acquired or inherited ownership of socially necessary instruments, such as railroads, steel mills, and factories, can work their will upon the rest of mankind. These essential instruments are social in their purpose and importance; they cannot safely be allowed to fall into the hands of individuals. Finally all social classes—wage earners, employers, capitalists, landlords, and others—are seen to be slaves to the passions of self-interest and acquisitiveness which the competitive system arouses.

The socialists believe that the order of society which they propose will be one of real freedom. Men will not work for the profit of other men but for the group as a whole and under its commands. The instruments of labor and the resources of nature will be owned and managed collectively under democratic government. The passions of greed and self-seeking will be held in check by arrangements which play upon better qualities in human nature.

Socialist analysis of capitalism: Karl Marx. The socialist analysis of the development and present tendencies of the capitalist system is derived in the main from the work of Karl Marx, who was the founder of almost all the modern schools of socialism. The *Communist Manifesto,* written in collaboration with Frederick Engels in 1848, and *Capital,* an analysis of the economic laws of capitalism and a forecast of its future development, which appeared in 1867, have profoundly influenced all subsequent radical movements. For our purposes the essential doctrines of Karl Marx may be presented under two heads, the class struggle and the trends of capitalist development.

The class struggle. The doctrine of the class struggle is a phase of the materialistic conception of history which is the keynote of Marxian historical philosophy. This interpretation of the historical process explains all institutional developments in terms of the underlying economic institutions. Government, religion, family system, art, literature, and morals are by-products of man's economic activities, harmonizing with the prevailing system of production

and exchange. The changes which have appeared in these institutions in the course of societal evolution have been the effects of prior changes in the economic basis of society. This doctrine, in some degree of comprehensiveness, is now believed to be true by most students.

Marx carried the doctrine one step further. To him not only were the economic institutions basic and determining, but these institutions had their own vital centre in the class struggle. Every system of production which mankind has hitherto devised has created a hierarchy of classes, exploiters and exploited. In classical society slave, plebeian, patrician, and noble, in the Middle Ages serf, vassal, knight, and feudal lord were ranged in an ascending order of increasing privilege.

Not only is class antagonism the essential characteristic of any historical social order at any given time, but this antagonism has produced the dynamic power which has caused society to evolve and change its structure. For the interests of the classes have clashed and forced them into continuous conflict. Revolutions in economic institutions, themselves the cause of all great changes in other institutions, have resulted from the class conflict when a previously exploited class gained power to throw off the yoke of the oppressor. Thus "the history of all hitherto existing society is the history of class struggle." But each such revolution merely re-established the system of exploitation in a new form, for the victorious class in turn organized a system of production with itself in the seat of power and the rest of mankind subservient to it.

The last great change of this type occurred when the bourgeoisie arose and overthrew their feudal masters, thus instituting the system of capitalism. Out of this grew the present class alignment, capitalists and laborers, and the present struggle between these two classes. The outcome of this struggle lies in the future, but it is predicted that the wage earners will eventually succeed, as did the present capitalists, in over-throwing their exploiters.

Their triumph will mark the end of the class struggle, since they will reorganize the industrial system so as to remove the possibility of exploitation. This will be true because the wage earners, being a vast majority of the whole population and committed to coöperative action, will organize the economic structure on collectivist lines, in which each will work for the welfare of the whole group. Thus

the socialist revolution will usher in a classless and therefore stable society.

Trends of capitalist development. As a result of his study of the capitalist society of his day Marx thought he had discovered the existence of two motive forces actively tending toward socialism and a third facilitating the eventual transition to socialism. The active forces were: (1) the tendency toward concentration and (2) the increasing number and misery of the laboring class; the facilitating factor in capitalism was (3) the recurrent collapse of the productive system in times of crisis.

Tendency toward concentration. As to the first of these forces, it is argued that capitalist development tends in two directions toward increasing concentration: first, the concentration of industry into larger and larger business units; secondly, the concentration of wealth and income in the hands of a diminishing number of people.

Industrial concentration is said to be due to the superior efficiency of large industries, which enables them to undersell and crush out smaller competitors. We must note to Marx's credit that this prediction was made before the trend toward large-scale production had set in. But Marx set no limit to the process, leaving his readers to infer that concentration would continue until a single gigantic business unit monopolized each of the major fields of industry, including agriculture. In agriculture the prediction has been falsified by subsequent events. How far it has been verified in other fields we have seen in another place.

This tendency was to lead to socialism, for the double reason that collective control would become both necessary and easy. It would be necessary in order that society might protect itself from the menace of these monopolies. It would be easy to establish because of the disappearance of effective competition, and easy to administer because of the relative simplicity of operating the few centralized productive units.

Concentration of wealth was inferred, partly from the preceding tendency, partly from other evidence. Marx made the mistake of assuming that concentration in industry of necessity involved concentration of ownership. He wrote before the day of the corporation, whose joint stock principle permits growth in the scale of the business unit coincident with an increase in the number of those who hold property rights in it. Great as are existing inequalities in

distribution, we have no evidence that they have grown greater since Marx's day or are now growing greater.

Increasing number and misery of the laboring class. The tendency toward concentration, coupled with other features of the capitalist system, was expected to produce an expanding propertyless class, with a steadily declining economic status. The crushing of small enterprises by the growing monopolies would precipitate former small capitalists into the laboring class. The growth of an industrial reserve army, the effects of periodic depression, and the absorption of an increasing fraction of income by the rich would cause increasing misery for this ever-growing group. The rich would grow richer and fewer, the poor more numerous and poorer.

This process would cause the system to break down through sheer inability of the masters "to maintain their slaves in their slavery." It would also cause the victory of the proletariat to put an end to the class struggle, since it would be a victory of "the immense majority in the interests of the immense majority."

The role of crises. Marx made much of the tendency of industrial society to fall periodically into crises and depressions. In his opinion crises were the inevitable outcome of the exploitation inherent in the capitalist system. The wage system resulted in withholding from industrial laborers a substantial fraction of the values attributable to their productive efforts.

This would be momentarily profitable for the employer, but for the system as a whole it would mean that goods were continuously produced in excess of the ability of the great mass of the people to take them off the market. This situation would run along for a time until the accumulated stores of unsalable goods precipitated a crash, followed by a depression, during which production would stop until the surplus of products had been liquidated. Marx believed that such paroxysms of the industrial system would not only recur periodically but would grow progressively more severe.

This theory of crises held a place in Marx's prediction of the inevitability of socialism. In the first place, each crisis and depression would play its part in increasing the number and the misery of the propertyless classes, thus promoting the trend described in the preceding paragraphs. Moreover as each collapse became more complete and recovery more delayed and difficult, the time would come when the recuperative powers of the system would prove inadequate to the task of restoring equilibrium.

The crisis itself was a challenge to the validity of the capitalist system, calling attention to faults which permitted an excess of the products of man's labor to exist coincidentally with widespread misery and want among the laborers. It therefore predisposed ever-increasing numbers of people to turn against the system and seek for a substitute. There would come at last a collapse so cataclysmic that revolt would occur before the protagonists of capitalism could bring about recovery. Both Marx and his immediate followers believed that the transition to socialism would come about during a business crisis.

General appraisal of Marxian doctrine. Although there are serious errors in Marx's analysis of capitalism and many of its specific predictions must be dismissed as unproved, there remains much of value to the student of social evolution. Marx's basic interpretation of history has provided a convenient and illuminating approach to an understanding of the historical process. His analysis and forecast of capitalist development, written as it was in the early stages of industrial society, displayed rare powers of insight.

But the Marxian doctrine is important chiefly for its influence in shaping the development and the present character of socialist thought. It placed socialism on a new and firmer basis. There is real validity in the central thesis; *i.e.*, that every thorough evolution in society's institutions must come about as a natural unfolding of forces at work in the preceding order.

Until the rise of communism this principle had great influence on the practical tactics of all the followers of Marx. It convinced them that capitalism must run its course before socialism is possible. The factory system in industry, with its corollary of a large wage-earning class, democracy in the political field, education for the mass of mankind, experience in collective action gained through labor organizations—these phases of capitalism must be nourished to maturity. Violent premature revolutions would fail disastrously.

Marxian socialists were forced by the logic of their system of thought into peaceful and legal forms of activity. Many of them believed, as did Marx himself, that the final overthrow of capitalism could not occur without violence and disorder, but their preliminary program was evolutionary rather than violent.

This outlook has continued to dominate the doctrines and tactics of all the principal socialist groups except the communists. These revolutionaries, although they claim Marx as their spiritual

leader and profess to derive their doctrines from him, are convinced that socialism can be established even in backward countries which have had no experience of capitalism. They also deride the peaceful political tactics of other groups and place reliance on direct action. So complete is this break in doctrine and strategy that it would be misleading to attach the term "socialism" to both wings of the movement. Accordingly we shall deal with them separately in the discussion of program and campaign which follows.

The socialist program. All socialist groups agree in that they propose to abolish private ownership in the means of production and to substitute for competition some form of authoritarian control over productive processes, but they differ as regards the form of collective ownership and control. The principal line of division is between (1) those who would make the present territorial state the owner of productive instruments and would vest control of economic activity in the chosen representatives of all the people and (2) those who would place each branch of industry under the government of the workers in the industry, relying upon various devices of federation to attain harmony within the economic system as a whole.

The syndicalists and guild socialists are examples of the latter branch of the socialist movement. Their ideas have attained some prominence in a few countries, but on the whole they are far less influential than are the representatives of the other wing of the movement, who are given the name "state socialists" from the pivotal feature of their reform program.

In the early stages of their development, the state socialists drew up detailed descriptions of the social structure which they proposed to install. The central government was to own the industries whose significance was nationwide; to the regional units and the municipalities would belong those which supplied only local markets. Some allowance was made for individual enterprises, provided they refrained from employing hired labor, and for coöperative enterprises owned by the workers. Although the land was to be nationalized, provision was made for the operation of farms by family units rather than solely by hired employees of the state. All these forms of non-collective enterprises were to be brought under control by strict regulation of the prices of their products, and by requiring them to adjust their productive activities to an over-all economic plan for the nation.

Since no form of investment of an income-producing type would exist, there would be no incentive for individual saving, beyond that created by the desire to own personal and household effects. The state was to be the only saver of productive capital. Since the people would, for the most part, find employment only in state-owned enterprises, wages were to be the predominant form of income, and the size of this income was to be determined by political authority. Even the incomes of the self-employed were to be similarly determined through control of prices.

Thus the central government, with its subordinate political units, was to be the dominant power in the lives of all. The political constitution of this projected socialist commonwealth provided that this great power should be brought under democratic control through universal suffrage, short-term election of all policy-making officials, proportional representation, and similar devices.

These blueprints of the new socialist order were not viewed as sacred and inviolable, for the state socialists have been more opportunistic than doctrinaire in their outlook. In all democratic countries, they formed local and national political parties, to press for government ownership and control wherever private ownership appeared to them to threaten popular welfare and competition appeared ineffective or destructive. They have been willing to leave as large a sphere for individual initiative as is consistent with the general principle that natural resouces and the larger aggregates of capital shall be used for the common good rather than for private gain.

The Socialist campaign: Tactics. In most democratic countries socialism as a practical political movement has taken on a moderate, temporizing character. Political parties are formed locally and nationally to agitate for an extension of government ownership and regulation of industry. The immediate objectives are usually limited: municipal ownership of the principal services, nationalization of railroads, mines, and public utilities, legislation to limit the power of employers and strengthen organized labor, and the like.

The peaceful methods employed and the relatively minor changes proposed at any one time often conceal the full implications of the movement. But this gradual encroachment on the domain of capitalism is expected eventually to produce an economic order so extensively socialized that the establishment of a thoroughgoing so-

cialist state will be easy and natural, and it is in terms of these final results that the whole movement must be judged.

The development of socialism in the separate countries is too complex a subject to be given adequate treatment here. Everywhere the movement is divided within itself over questions of doctrine and tactics into factions, which in most countries are organized for political purposes in opposition to one another; this quite apart from the separate organization of the communists, who are bitter antagonists of the socialist movement as a whole.

The divisions within the socialist movement proper usually reflect conflicts over the relative emphasis to be placed on organized labor, on the one hand, and the territorial representative state, on the other, as the sovereign body in the socialist economy, or differences of opinion as to the rapidity with which modern industrial society should be socialized. Without going into detail it can be said that the Social Democratic Party, to be found in many countries, the Labour Party of England, and the Socialist Party of the United States, usually representing the majority wing of the movement, stand for reliance on political government and a cautious experimental rate of change, while the syndicalists and the radical socialists favor revolutionary action through labor unions and a hasty continuous drive toward socialization.

Political action. In most industrial countries, excepting the United States, socialism has been a powerful factor in the political and social life. The dictatorship countries, whether communist or fascist, crushed out a once strong socialist movement within their borders. The socialists however have controlled the governments of many municipalities throughout Europe and in Great Britain. They always form a strong opposition party in the national assemblies and at times have controlled the central governments in European democratic states.

Typical of conservative socialism is the Labour Party of Great Britain, formed early in this century as the political arm of British organized labor. From small beginnings this party grew in twenty years to become the strongest single political group in the country. In 1924 and again in 1928 the Labour Party, though still in the minority, was called upon as the strongest single group in Parliament to form the ministry and conduct the affairs of the nation. Because of its insecure tenure, the party did not at those times attempt to put the ideas of socialism into practice on a national scale.

A general election in 1945 swept the party into power on a thorough-going socialist platform, which the Labour Government has been proceeding gradually to put into effect.

Although the socialist movement in America has been in existence for many decades, it has never developed into a strong national party. The prevailing individualism of the people and the relatively high standard of living produced by a system of free enterprise operating in a naturally rich country have been deterring factors. Furthermore the peculiar constitutional structure of the country places obstacles in the path of minority third parties contesting for control of the national government. In local elections the socialists have at times gained control of municipal governments. They have also captured seats in the state legislatures and elected some members of the House of Representatives. But these successes have not enabled them to reform the national institutions along socialist lines.

The American socialists, like those of Britain, have eschewed violent tactics and pursued legal methods of political activity. They have taken part in national elections since 1904, but with small result in terms of popular vote and with no consistent increase in their following.

It would be misleading, however, to measure the strength of the socialist movement by the record of the Socialist party. A great many people who accept the socialist doctrines of reform do not belong to the party, preferring to work within the major political parties. Many programs, such as social legislation, governmental regulation of business, and direct government activity in certain fields of economic enterprise—originally sponsored by the socialists—have come to be adopted and put into effect by the major parties. There is in fact little difference to be found today between the economic policies of the radical wings of the major parties and the immediate programs adopted by the state socialists as steps leading gradually to complete social reconstruction.

Appraisal of socialism. Before proceeding to a discussion of communism, it may be profitable to return to the socialist indictment of capitalism outlined earlier in this chapter. Every thoughtful person will admit that each of the charges in this indictment contains elements of truth. The capitalist, or free enterprise, system has many and serious faults—faults of inefficiency, of injustice, of retarded adjustment to the changing conditions of human life. The socialist

movement is to be credited with a great service in having focused attention on these faults and accelerated their correction.

Since socialism has nowhere been in peration long enough to demonstrate its superiority or inferiority to capitalism (the communism of Soviet Russia will be given separate consideration), attempts to appraise its prospects must deal with its doctrines and programs. On this basis the case in favor of socialism is not convincing.

In the first place, the socialist indictment of capitalism is overdrawn. It both exaggerates the inefficiency of capitalism and ignores the ability of the system to correct its own faults. This type of economic order has demonstrated its ability to maintain large populations on the highest average level of comfort ever known, to provide an unprecedented supply of capital instruments, constantly increasing their quantity and improving their quality, to supply the individual consumer day by day with goods drawn from all the resources of the earth. Furthermore, capitalism is a flexible system, capable of modifying its basic institutions without destroying their fundamental utility.

Thus, collective bargaining can be substituted for the individual contract when the latter works injury to the weaker members of society. Social insurance can be developed to safeguard the workman against many of the hazards of the wage earner's life. Labor legislation can assure him minimum standards of safety, sanitation, even of income. Collective provision may be made for the services which private initiative will not supply, such as education, agencies of public health, recreation facilities. Producers can be restrained from undermining and destroying each other by unfair competitive practices. The consumer can be protected against adulterated goods and scrimped quantities by laws enforcing standards of purity and honesty of weight and measure. The overweening power of great business can be restrained by public control.

All these things can be done without discarding the basic institutions of society. Although we may not be content with the rate of progress attained along these lines in the past, that there has been progress no one can deny. The true conservatives of our day are those who lend their support to these reforms; those who oppose timely readjustment of traditional ways and methods must be ranked among the enemies of the present order.

In attempting to improve upon the system of free enterprise

which it would supplant, socialism will encounter certain difficulties inherent in its own program. First, there is the general problem of human motivation in economic affairs. The free enterprise system has attained its great productiveness by calling into play two of the most powerful of human incentives: acquisitiveness, and the desire to excel in competition. The institution of private property, with its concomitant, the right of inheritance, provides for the former of these incentives; individual enterprise, the latter.

Socialism proposes to suppress the motive of acquisition almost entirely and to limit severely the material reward to be attained through success in competition. In place of these incentives, after providing general guarantees of security, socialism will rely upon motives of social service and reward in the form of political and social preferment for work well done. Human nature may respond to such incentives and rewards, but there is grave reason to doubt that they will prove as effective in stimulating economic activity as the motives which have been responsible for the material success of capitalism.

Apart from problems of human motivation, socialism will encounter grave practical difficulties. All forms of socialism impose on some public body the task of determining the prices of different goods and services. Is not this a task too great for human intelligence in so complex a society as ours? Price fixing involves a determination of quantities to be produced and a rationing of individual consumption. After wages have been fixed for all classes of labor, the state or some other public body must then see to it that society is supplied with the right amounts of each type of worker. When a price has been set upon wheat or coal or shoes, some authority must then supervise production and consumption. These difficulties multiply as one ponders the problem of authoritative control.

Again some method must be worked out to select the natural leaders of economic enterprise, to advance them to the positions for which they are best qualified, and to give them the requisite authority to make their powers effective. Each of these steps presents serious difficulties. Under capitalism, selection of industrial leaders results in the main from demonstrated capacity to excel in open competition, and those who succeed are given broad powers to control and direct industrial processes. This natural process of trial and error is, it is true, frequently blurred by the inherited privileges of

mediocre men, but in the vast range of economic enterprise these faults of the system are exceptional.

There is grave doubt that socialism can devise any method that is not inferior in promoting efficiency. Certainly the methods proposed in their program—selection by civil service examination, or by popular acclaim, or by political preferment—are not reassuring. Such methods have always worked against the innovator in favor of the man who conforms to traditionalism, so dear to the hearts of public officials. These methods of selection tend also to weaken discipline and circumscribe the initiative of managers, by making tenure in office depend on the good will of subordinates.

Fourthly socialism must provide for the maintenance and improvement of the capital equipment. No possibility would remain for the saving of capital by individual initiative, since interest would be destroyed as a private income. The state, as recipient of the social income, can withhold a portion to be reinvested, thus forcing the people to save against their will or without their knowledge. But in view of the impatience of men for immediate income and of the subservience of officeholders to popular clamor, is it likely that any group of public officials could be found with sufficient courage and foresight to administer such a system wisely?

As regards the improvement of technical processes, it is argued that these come in final analysis from scientists, who labor without thought of material reward for the sake of their love of truth. This may be granted without removing the difficulty. Today the discoveries of science are adapted to the economic service of mankind, not by the discoverers, but by other men who are certainly intent on personal reward. Pure science might possibly flourish under socialism, but invention would lag behind because of the apathy, the obtuseness, or the abhorrence of change of public officials.

Communism: Its rise in Russia. Before the First World War, a socialist movement based on the doctrines of Karl Marx had made some headway in Russia. This movement had split into two factions, one of which, the Bolshevists, was destined to lead the Russian revolution and to establish the Communist Party as the sovereign power of the state.

What distinguished the Bolshevists from the rival socialist faction was their repudiation of the Marxian dogma that socialism can be established successfully only in countries with mature capitalistic institutions—a condition which Russia was obviously unable to ful-

fill. The Bolshevists believed that socialism could be set up in backward countries, if established by direct action at a time of economic or political crisis and maintained by a ruthless dictatorship using terroristic methods.

Under Lenin's leadership, the Bolshevik leaders took advantage of the chaos resulting from Russia's defeat in the First World War to carry through a successful *coup d'état,* and when settled in power adopted the name, Communist party. Discarding their affiliation with the social-democratic movements of other countries, the Communists became bitter opponents of the type of socialism which we have thus far discussed. They have organized opposing parties under their own name throughout the world, and today the conflict between communism and socialism is everywhere as bitter as the clash of either of these groups with their common conservative opponents.

The original communist program. The Soviet Government of the present day does not practise communism in its original purity, having been obliged by international and internal conditions to compromise with many of the basic principles. It is necessary, however, to examine the doctrine of pure communism in order to understand the ideal society envisaged by this movement.

One distinguishing feature of communist doctrine in contrast with socialism is its conception of revolution. Not only do the communists reject the Marxian dogma that socialism can be established only in countries with mature capitalistic institutions, but in all countries they are scornful of the peaceful tactics of the orthodox socialists, believing that dispossession of the capitalist classes can never be brought about by these means, while to adopt them merely enervates the movement and commits it to a policy of flabby compromise.

The communists base their strategy squarely on the class war, a succession of armed clashes with the forces of capitalism, culminating in a violent revolution on the Russian model to be consolidated by a reign of terror. In harmony with this outlook they repudiate democracy, both in the organization of their party and in that of the revolutionary state, and propose a dictatorship of self-appointed leaders commanding a loyal well-disciplined proletarian army.

The essential features of the ideal communist society must be inferred from the theoretical literature of the movement. They

may be summarized as follows: There is to be collective ownership of all forms of industrial enterprise, administered by the government; labor alone is to give right to income, and all labor is to be performed for the government as the only employer; the entire social income is to accrue, in the first instance, to the government, to be distributed according to communist principles.

The method of distribution differs from that of other socialist schemes. Wages are to be the same for all workers regardless of function and are to be paid in labor tickets, which are redeemable in goods out of the general fund. Money is to disappear, and the only form of exchange will be that of "service for counter-service"; that is, labor service of the individual for the state and counter-service by the state in the form of consumption goods, supplied to the individual. Many important types of goods—such as transportation, housing, light, fuel, perhaps food—are to be free for all citizens. Equality is to be the rule, and any departure from the rule is to be justified only on the ground of the greater need of the favored individuals. "From each according to his ability" is the formula governing the social service of the individual, and "to each according to his need," the formula determining his reward.

The transition stage of the revolution is counted on to give the dictators opportunity, not only to bring the country's social institutions into conformity with this ultimate purpose, but to remould the behavior of the people and give them a motivation appropriate to a communistic society. When these ends have been achieved, dictatorship is to give way to a completely democratic system of government. Indeed the state itself, as a coercive organization, is to wither away and disappear as the people become motivated toward voluntary coöperation. The term "government," as used by the communists, connotes merely the executive organization of such a freely coöperating group.

Communist organization: Within Russia. The Communist Party of Russia provides the prototype of the movement's organization in all countries. It is to the organization of the Communist Party itself, for its own government, to which we refer, and not the structure of the government of the Soviet Union. Its essential features are: complete centralization of authority in the hands of the party's leaders, provision for fairly liberal debate of policy and tactics up to the time when decision is taken as to the "party line," ruthless suppression of all differences and the requirement of undeviating

loyalty to the leaders when once decision has been reached. In other words, there is avowed dictatorship within the party, as within the political state, which the party controls.

Membership in the party is not a thing to be undertaken lightly. To join requires a confession of faith and devotion, a willingness to submit to arduous testing during a period of probation, and a readiness after becoming a full member to bear heavy financial and labor costs on behalf of the party's program. Members are taxed heavily in proportion to their earnings and are expected to give all their spare time to party works under the orders of their officers. Periodically the membership is "purged," through a comprehensive inquisition, which eliminates those of faltering faith and lukewarm works. Membership, in other words, is like that of a strict religious order, requiring zeal, devotion, and sacrifice. So it is in the parties of other countries.

For this reason the Communist Party of Russia has never been numerically large, never including more than a trifling percentage of the people. But what is lost in numbers is more than made up by the discipline and drive of the movement. This quasi-religious aspect of communism, with the irrational and uncompromising attitudes which it evokes, is a salient feature of the movement in all countries.

The structure of the party is based on a nucleus or "cell," a local unit formed, in a small geographical area, an industry, or other organization, by three or more members. These local units are grouped into convenient regional organizations; these in turn form larger territorial groupings, which are brought together into the national party government at the top. Disregarding complexities of detail, it may be said that the structure is erected by electing delegates from the lower units to the higher, culminating in the election by the largest territorial subdivisions of delegates to the All-Russian Congress, reputedly the supreme authority of the party.

But the actual lines of authority are concealed by this ostensibly democratic procedure. The Congress is an unwieldly body of large membership, meeting for short periods at long intervals. It elects a Central Committee to meet more frequently on matters of policy, a Central Control Commission to serve as a supreme disciplinary authority over the party membership, and a Central Revision Committee to audit the party finances. The Central Control Commission, also a large body, delegates its authority to an elected execu-

tive board known as the Presidium. The Presidium, the Central Committee, and the Revision Commission join to elect the general secretary of the party and the Political Bureau, a small body in continuous session composed of the key men in the various central organizations.

Through this complex hierarchy of interlocking organs supreme authority is gathered into the hands of the secretary and the Political Bureau. The secretary controls the paid party workers in all the local units and, using them as a personal following, becomes the leader of the entire party. It was by means of this position that Stalin first rose to dominance, and he subsequently achieved his present position of absolute personal dictatorship by using this machine to crush any of his former chiefs and colleagues among the older Bolsheviks who challenged his authority or differed with his policies. The Political Bureau is now composed of Stalin's closest associates. It is in reality the sovereign body in party government under the leadership of the Secretary, since it exercises all the powers of the Central Committee, exerts party discipline through the Control Commission, prepares and steers the program of the Congress, and controls the entire press and other publicity organs of the country.

International organization. So much for the national setup in Russia. The same group of organizations which elects the secretary and the Political Bureau also appoints the delegates of the Russian Communist Party to the Third International, which attempts to impose centralized control over the policies and tactics of the party's branches in other countries. Communism in its pure form is an international creed, ignoring national frontiers; the party is a single organization throughout the world, owing common allegiance to a single leadership in the Third International, which is dominated by the Russian branch only because of the latter's numerical superiority.

The International is supposed by its constitution to meet every year, but following a few annual meetings immediately after its origin in 1919, the sessions became less and less frequent, and none has been held since 1935. In this infrequency of meetings, as well as in the alterations of its announced policy, the International has reflected the political expediency of Soviet Russia, particularly the growing desire of that country to avoid arousing the antagonism of other nations. For a time, beginning during the Second World War

and extending until the fall of 1947, it was the avowed purpose of the Russian leaders to allow the International to die and thus to subordinate the emphasis on world-wide revolution which is an essential feature of the original communist program, but these intentions were abandoned as international tension developed during the postwar period, and an organization of the Communist parties of seven European countries was formed.

The pronouncements of the Third International, as the oracle of communism, have been concerned far more with strategy than with doctrine. The basic doctrines of the movement are held in veneration as an inspired creed, revealed in its unalterable essentials by Marx and interpreted by later disciples, principally Lenin and Stalin. It is assumed, somewhat contrary to fact, that these present-day interpretations are merely an exegesis, not an amendment, of Marx, and it is also taken for granted that the creed thus interpreted has the unquestioning adherence of all loyal members everywhere. The International took pains from time to time to expound to the world movement these interpretations of basic doctrine by the Soviet leaders, to justify by reference to the creed the current policies of the Russian government, and thus to obtain the unqualified support of world communism for the regime in Russia.

Since unanimity in the matter of doctrine was assumed, the chief task of the International became that of issuing instructions on immediate tactics to the foreign branches of the party. In this it showed itself to be dominated by the current needs of the Soviet Union. Its early meetings turned on the theme of class war and world-wide revolution; all branches of the party were encouraged to put themselves on a war footing, holding themselves aloof from all other movements, however radical, and fomenting conflict by guerilla tactics wherever opportunity appeared.

As Russia became more and more absorbed in internal reconstruction, requiring stable commercial and political relations with other countries, and especially as the growth of fascism brought its dangers of attack upon the Soviet Union and of suppression of communism in other countries, this line of strategy was altered. The orders issued by the Seventh Congress in 1935 disclosed a complete change of front as contrasted with the early days. The communists of all nations were ordered to drop their bellicose tactics and to cooperate with, and if possible to join, every radical and progressive reform movement in their respective countries.

Dimitrov, head of the International, defined this as "the Trojan horse policy." It was justified on the score that the Communists must make common cause with the liberals against the growing power of the Fascists which is destructive to both. To be sure, communism as a method of government has less in harmony with liberalism than with fascism, but the latter had issued a death sentence upon communism, while the "liberal" and progressive modern democracies, in accordance with their principles, were tolerant of it. Reference to these tactics as a Trojan horse policy disclosed their ultimate objective. By "boring from within" the liberal organizations of society, it was hoped eventually to capture these organizations and turn them to the long-run purposes of communism.

There is clear evidence that the formal disappearance of the Third International did not weaken the world-wide unity of the communist movement nor remove the hegemony of the Russian party. Throughout the events leading up to the Second World War, during the war itself, and in the postwar developments, the communist groups in other countries have consistently attempted to mold the policies of their own governments in harmony with those of the Soviet Union. Abrupt changes in the international outlook of Russia resulted immediately in equally abrupt reversal of communist policy elsewhere.

During the Second World War, the communist groups in other countries, attaining respectability from the alliance of Russia with the democratic nations, re-established their independent parties, while continuing their tactics of boring from within. After the defeat of the Axis Powers, the strength of the movement grew rapidly in the western European nations, under the favoring influence of the prevailing conditions of distress and disillusionment. In the Baltic states and among the peoples of eastern Europe, the spread of communism has been fostered by the influence of the Soviet Union, under whose shadow these countries have lain. Postwar elections in western Europe would indicate that socialism rather than communism will ultimately prevail in this region, but it is too early to predict to what extent communism will spread as the result of the war and its aftermath.

Policies in the Soviet Union: Recent history. In dealing with the Russian experiment as an example of communist economics, we must disregard the experiences of the years immediately following the revolution, for this period was too greatly disturbed by civil

war and foreign invasion and by the profound economic collapse of the country to permit the communist leaders to concentrate on their program of social reconstruction. In the following years, however, and particularly since Stalin solidified his power, the dictators have held complete command of the situation, both within their own party and throughout the country at large, and the economic policies they have adopted during this time may fairly be taken as representative of communism in its early stages.

A turning point of policy was reached when Stalin and his colleagues decided that their experiment could not wait upon successful class revolution in other countries but must proceed to establish socialism in Russia in isolation from the capitalist structure of other modern nations. To some extent the Russian program was colored, during the 1930's, by the dictators' fear of attack from without. This fear resulted in a stress upon national self-sufficiency, the stimulation of patriotism, in disharmony with the theoretical internationalism of the communist creed, and the large investment of capital and labor in military preparedness, which has retarded elevation of the standard of living of the Russian people. During the Second World War, of course, the entire energies of the country were devoted to military ends, and the international uncertainties following the war prolonged this phase of communist development.

Planned economy. The essence of the soviet program is a comprehensive economic plan, which embraces every phase of the national life and attempts to carry the entire social order toward predetermined objectives. Certain phases of the plan are noneconomic in character, dealing with such problems as the elimination of illiteracy, the promotion of education, hygiene, and public health, cultural activities, and the like. These are among the most interesting aspects of the plan, both in conception and achievement, but we must disregard them here for lack of space.

The planning mechanism is the State Planning Commission, created by the central government, together with its subordinate bodies for regional divisions and separate economic enterprises. The commission is composed of experts in the different departments of the national economy, dominated by party appointees who hold the program true to the "party line." Its function is to prepare detailed schedules for every line of industrial and agrarian activity, covering the kinds and quantities of goods to be produced, together with the projected costs and prices, the rates of capital increase in various

industries, the increase in the number and technical efficiency of laborers, and the like—the whole plan having relation to a term of years immediately following its adoption.

During the operation of the plan, the Economic Council of the central government exercises supervision over it, checking current rates of production against the scheduled forecasts, and altering the schedules from time to time in accordance with changing political and economic conditions; the whole weight of governmental authority is behind the council's decisions. Planned economy in Soviet Russia does not mean rigid, hard-and-fast adherence to the details of a program, but a flexible policy which, though outlined in advance and moving steadily toward predetermined general objectives, allows for considerable readjustment along the way.

Planned economy really dates from the adoption of the first "Five-Year Plan" in 1928. A second "Five-Year Plan" followed in 1933, and on January 1, 1938, the third "Five-Year Plan" was launched, but this was set aside with the outbreak of war. With the cessation of hostilities, another plan was adopted, looking toward the repair of war damage and the rehabilitation of peactime industries.

The principal features of these plans may be summarized as follows. First was a determination to socialize the entire industry of the country under government ownership and operation, extinguishing private enterprise in all branches of industrial production and distribution. Second was the intent to establish socialism in an agrarian economy dominated at the time by millions of independent small-scale farms. Third was the plan to expand on a gigantic scale the industrial capital equipment of the country, to change its technique from small-scale to mass production and from handicraft to power-driven machinery, and greatly to increase its rate of output and lower its costs. Similarly, the aim was to increase the output of foodstuffs and raw products of the sort required to make Russia industrially self-sufficient.

These were the major objectives. As will be seen, they involved a fundamental revolution in economic organization, as well as an increase in productivity. The intention to raise the standard of living in the country by increasing the quantity and variety of consumable goods may be mentioned as a fourth objective. But the drive toward military preparedness, with its emphasis on the machine-making

and war equipment industries, caused this phase of the plans to be set aside more or less completely.

Soviet economic organization. Almost every department of the nation's economic life has been successfully transformed into a system of state socialism. The government owns and operates the railroads and other public utilities, the mines and forests, and all the factories; it monopolizes trade, both domestic and foreign; the banking and credit mechanism is in its hands.

Agriculture is to some extent an exception. Under the soviet constitution, legal title to all lands is vested in the state, and in accordance with this principle there are vast farms owned and managed by the government. But the attempt to force the state farm system into effect throughout the country proved a disastrous failure because of the passive resistance of the peasants.

The government was compelled to adopt a modified form of socialization. After each peasant family had been given outright control of a small area of land and ownership of some livestock and farm implements, the remaining land was brought into a system of producers' coöperation, in which numerous individual peasants merged their land holdings into a single productive unit, which they managed collectively under governmental supervision.

There is no private employment of wage labor on the individual farms; the socialized industries provide the only source of wages. Small household industries are allowed to exist, provided they conform with the law prohibiting the hiring of labor and sell their products at prices fixed by the government.

These vestiges of private industry are recognized as anomalous, their existence being justified by the government's inability thus far to provide an adequate supply of consumption goods through its own industries. At the outbreak of the Second World War the reconstruction of the country's economy along socialist lines was virtually complete.

Growth of soviet industry. From rudimentary beginnings, soviet industry developed rapidly under the impetus of planned economy. An impartial review[1] of the situation in 1938 disclosed that during the preceding decade industrial production had increased eightfold. Especially rapid strides had been made in the heavy-metal and machine-making industries, which were virtually nonexistent at the

[1] See *Foreign Policy Report*, June 1, 1938, published by The Foreign Policy Association, New York.

time of the revolution. Russian-made agricultural machinery, factory equipment, airplanes, automotive machines, for which the country had formerly relied on foreign supplies, were produced in notable quantities. Electrification was proceeding at a rapid pace. The production of industrial raw materials had increased nearly fourfold during the ten-year period.

Toward the end of this period, Russia had begun to focus attention on military preparedness, a course which she followed with increasing intensity during the years immediately after 1938 and of course during the war years. Her impressive war record, though aided by substantial quantities of military equipment from her allies, testified to the success of her industrialization program.

The program was not free of failures and shortcomings. A lack of trained managers and skilled workers retarded production along many lines and resulted in many defective goods—especially in the early years of the plan, when a considerable part of the output of the new machine industries proved so inferior in quality as to be well-nigh useless. These shortcomings were dramatized, in typical communist fashion, by the summary execution of managers held responsible for them. But techniques of production improved with experience, and a broad training program for workers in the different trades and for managers had, by the outbreak of war, begun to provide more adequate supplies of the human resources indispensable to any industrial society.

The weakest point in the record was the failure of the planned economy to provide for the everyday needs of the people. Articles of consumption remained very scarce and of poor quality; in fact many articles of household and personal use, which in capitalist countries are possessed by even the lowest income classes, were so nearly nonexistent as to be classified as rare luxuries. Thus the plane of living in Russia remained almost unbelievably low. This fact however does not necessarily argue the inability of an authoritarian socialist regime to provide a comfortable living for its people, but merely indicates that the dictators had intentionally given priority to the production of nonconsumable capital instruments, requiring the people to bear the sacrifices inherent in that policy.

Labor and wages in Soviet Russia. The communist program is conceived in the interest of the wage earner, who is to be set free from the slavery which is his lot in capitalist society, given an independence which he has never formerly enjoyed, and accorded a

larger share in the social income. In these as in other respects, the experiment in Soviet Russia has failed thus far to conform to communist principles.

Labor policy has had an interesting development. At an early stage substantially equal wages were paid to all workers regardless of function, the higher types of expert and managerial labor being placed on about the same level of reward as the manual workers. This scheme broke down completely and had to be abandoned, owing to its failure to stimulate the industry and efficiency of the abler workers. As the drive for rapid increase of productivity became more severe, the last vestiges of freedom and equality among wage earners were swept away. At the present time, the wage system differs very little from that which obtains in capitalistic industry, except that the average real wage is very much lower and the extremes of inequality not so great. Piece-rate wages, payment in proportion to product, and higher rewards for services which are scarce are all characteristics of the soviet wage system.

In like manner other socialist theories of labor management have had to be discarded. At first the powers of management were severely limited, and the workers were given virtual control over the operation of their respective industries. This system of industrial relations resulted in such hopeless confusion and inefficiency that the soviet authorities were obliged to re-establish managerial autocracy in industry. In fact the control and discipline of labor in Russia is now much more complete and dictatorial than in capitalistic nations.

It is true that labor is organized comprehensively into trade unions, but these organizations are entirely subservient to the state and, as advocates of labor's welfare on the job, are much less powerful than in the United States. Strikes are outlawed as treason against the state. Even unintentional inefficiency in union management is severely punished, as in 1937, when the entire higher officialdom of the Central Trade Council was dismissed from office, and four of the most prominent labor leaders were arrested. The unions no longer pretend to exercise any bargaining power in matters of wages, hours of labor, rates of output, or discipline. They are still allowed to function in adjusting minor local grievances, but on the whole they are used by the government, to strengthen its drive for greater productivity, and as agencies for administering the social insurance benefits.

The Russian people have always lived on a plane of poverty. It may be that the average real income is somewhat higher now than formerly, but it is still extremely low as judged by any standards known to modern industrial nations. Nominal wages signify little in terms of real income Prices of most consumers' goods remained so high that the wage income for the average family was sufficient only to cover the bare necessities of life.

In appraising the material welfare of the workers under the soviet regime, account must be taken of certain other benefits accorded them by the government. They are given security against unemployment, provided of course that they are willing to accept the terms offered by the only employer, the state. A comprehensive system of social legislation indemnifies them in case of accident or sickness. Maternity and child care, free education with stipends for students in the higher branches, and a certain amount of free medical service are also provided. For 1938 the government budgeted a total expenditure of over a billion dollars for these free benefits to the workers.

Soviet finance. In the final analysis the costs of such planned economy take the form of sacrifices by the people of freedom and material comfort. This does not refer to the large-scale destruction of life, the exile or imprisonment of great numbers of people, the forced labor of political convicts, all incidents of a ruthless dictatorship. What is meant is that in its routine economic operation the program has fed on the toll taken from the standard of life of the common citizen.

During the years prior to the Second World War, Russia was virtually unable to increase her resources by borrowing abroad. Yet she imported vast quantities of capital equipment and raw materials and drew other large quantities from her internal resources. In the final analysis this all came from the consumable incomes of the people, thus increasing the prevailing poverty. The government's power to determine the quantities of goods and services allowed to enter into direct consumption and the quantities to be employed in capital creation has forced the people to save, either against their will or without their knowledge.

But the day-by-day operation of the program has had to be financed, and the development of financing methods has produced some interesting results. In the first place, the Soviet authorities soon had to abandon the communist doctrine that money is un-

necessary in a socialist economy where the state monopolizes production and fixes wages and prices.

The money system which has been set up in Russia is ostensibly on a gold basis, with a standard unit—the ruble—which represents a fixed weight of gold. But this standard monetary unit is entirely hypothetical and does not exist in reality, being used merely as a unit of account for measuring values in government records and recording transactions in foreign trade. The money in circulation is irredeemable paper, whose value is determined solely by its purchasing power within the country.

This value is not governed by the principles discussed in our earlier study of monetary theory, for the reason that prices of goods have been fixed by law, and at times a rationing system has determined the distribution of goods among the consumers. In such a system there is no necessary connection between the quantity of money and the general price level. A given money income may fall into two parts: a portion which can be exchanged for a quantity of goods determined by authority; a remaining portion which has little value since there is scarcely anything to be bought with it.

The development of the program has produced a situation in governmental financing which is illuminating as a prototype of a socialist economy. By 1937 the government had gathered into its own hands 95 per cent of the national income as compared with 27 per cent in 1928. This of course refers to "income paid out"; a considerable amount of real income still accrues directly to the people, particularly the peasants, in the form of products from small land holdings and the like. This large share of the national money income accrues, in the first instance, to the state by virtue of its monopoly of all enterprises producing for the market.

Public finance under these conditions is chiefly an accounting of the receipts and expenditures of productive enterprise. The government's budget records some income from taxes levied directly on the citizens, but in the main its receipts consist of portions of the income of the government monopolies reserved for purposes of the state. These are recorded in the budget as taxes: a "turnover tax" on sales, a "profit tax" on railways, factories, farms, etc. But in essence all these exactions on the "profit" of business enterprise are subtractions from the incomes of the people. The "profit" is created arbitrarily by the power of the government to fix both the money costs and the sale prices of industry. What soviet public finance

means in simple language is that the dictators fix the money wages, which determine the people's income, sufficiently below the prices of goods to leave a large surplus for their own use.

Fascism: Present status. The defeat of the Axis Powers and the destruction of the fascist societies which they created have left many people with the belief that fascism is now a matter of historical interest only. This belief is erroneous. Even today fascist and quasi-fascist regimes hold power in certain countries. Furthermore most democratic nations contain organized groups professing fascist doctrines and attempting to propagate them among the people. Since fascism is still a living thing contending with other social philosophies for mastery of society, it is important that its essential characteristics be understood.

Comparison of fascism and communism: Similarities. It is a common opinion that fascism is the direct opposite of communism, representing ultraconservative principles. This opinion is supported by the fact that in Italy and Germany, where fascism triumphed, the movement arose as a belligerent opponent of communism. Regimes still surviving which are fascist in spirit also condemn all forms of radicalism, while in democratic countries where fascist ideas are allowed expression, the propaganda of the movement is likely to be supported by die-hard reactionaries in politics and industry.

But these aspects of fascism should not be allowed to conceal the fact that fascism and communism, in many essential respects, are very much alike. Both are revolutionary movements opposed to democracy; both adopt the same technique of revolution—direct action by a firmly disciplined minority under dictatorial leadership—and both aim to establish the same kind of authoritarian state practicing terroristic methods; both are destructive of free enterprise by reliance on a program of planned economy enforced on the people by dictatorial government. They are equally unacceptable to men of liberal mind, because both are alike destructive of human freedom.

Differences between fascism and communism. These similarities of communism and fascism have to do with techniques of organization and of operation rather than with ultimate purposes and objectives of the two movements. When they existed in actual operation in neighboring states, the two systems disclosed striking differences in these latter regards.

In speaking of these distinctive features of fascism which stand in contrast with communism, we must base our judgment on the only examples of the system which we have seen in thoroughgoing operation—the fascism of Italy and Germany. It may be that these features were due to some extent to fortuitous circumstances, such as the personalities of the leaders and the contemporary situations of the countries internally and in international relations; other fascist movements elsewhere may not include them all.

Judging however from what we have learned of fascism in operation, it is unlike communism in the following respects: First there is a difference in the concept of the state. The fascism of Italy and Germany worshipped the state as a spiritual entity, apart from and superior to the people. As a corollary, fascism is militantly nationalistic. Communism, on the other hand, proposes eventually to do away with the state and is in theory an international movement not bound by political frontiers. That Soviet Russia has turned to nationalism is explained by communist leaders as a necessary reaction to the condition of world affairs, admittedly contrary to their creed.

Second there is a contrast between the principles on which the dictatorship is based. A fascist state is to be dominated permanently by a small party, to which only the elite can belong, while communism, whose governing party is similarly organized in the early stages of the state, avows the intention of broadening its base until it includes everyone.

Third the two movements differ in their conception of social classification. Communism not only believes in the class struggle as a historical fact, but fosters it as a means of social evolution. Fascism, on the other hand, denies the historical necessity of the class struggle, explaining it as the artificial creation of demagoguery, and insisting that both proletariat and bourgeoisie receive their status from membership in the state and are bound by the same obligations of loyalty and service.

Finally one finds a number of minor differences which it is difficult to reduce to formula. For example, communism recognizes the essential equality of human beings, regardless of sex or race, granting to all the same legal rights and admitting them all to the same vocations, in the party, in the government, and in industry. Fascism, on the other hand, either excludes women from political

and social activity or subordinates them to men, and it has professed belief in extreme theories of racial superiority.

There is a difference too between the attitudes of the two movements toward religion. Every loyal Communist must be a professing atheist. Although the Soviet government permits freedom of worship, it limits the independence and authority of the church and looks benevolently on the efforts of the party to stamp out religious belief by propaganda. Fascism requires no personal opposition to religion of its adherents, and it interferes with the churches only when they, as organizations, are hostile to its purposes.

We shall not undertake a detailed study of the structure and operations of the fascist systems which have vanished as a result of defeat in war. Mention might be made, however, of two points in which the economic programs of the fascists differed from that of communism.

Fascism made no great use of government ownership and operation to put into effect the economic plan, but relied in the main upon private ownership operating under strict governmental control. This fact has led to the belief that fascism stands as a guardian of private property against expropriation by the radicals.

On the contrary, the truth is that private property lost all its essential rights under fascism, and private initiative ceased to have any meaning; this must inevitably be so in such a social order. Business enterprises produced in kind and quantity what they were ordered to produce by the state; their costs and sale prices were mostly fixed by decree; the internal relations of management with workers were forced to conform to patterns designed by political officials.

This difference in economic organization caused the financial systems of the fascist and communist regimes to differ. We have seen that in Russia government monopoly of productive enterprise causes the entire marketable produce of the country to accrue in the first instance to the state treasury, and that its monopoly position enables the government to employ these receipts in financing its program. The fascist governments, having no such resource, were obliged to rely on taxes and loans as means of appropriating the wealth and incomes of their subjects. This difference of procedure, however, means little in terms of human welfare when it is remembered that the fiscal powers of fascist governments are not limited by any guarantees of individual or property rights and are wielded by an officialdom unresponsive to popular will.

General appraisal of radical movements: Authority vs liberty. The complexity of similarities and contrasts which characterizes the radical movements of socialism, communism, and fascism makes it difficult to draw any general conclusions which have final validity. In the discussion which follows we shall emphasize their common attributes and especially those which stand in most definite contradiction to the essential features of a capitalist order.

One of the chief charges in the indictment of capitalism is that this system of society assures no real freedom for the average man. Granting this accusation all the weight it deserves, it certainly remains true that the evil has been magnified in the systems which have been substituted for the capitalist order. We need not dwell upon the incidents of massacre, imprisonment, and brutality which have horrified the democratic world; these are characteristic of dictatorship regardless of its social philosophy. But it is of the essence of capitalism, with all its inequalities in the distribution of wealth, power, and privilege, to rely upon the voluntary actions of individuals to provide for social needs. And it is equally essential to any controlled economy to substitute compulsion for voluntary action. The modern world now has experience of both methods in operation simultaneously and can easily judge of their results in terms of human freedom.

The constitutional socialists will hold this criticism inapplicable to their system, since they support democracy in industry and in political government. But it is questionable whether in practice they could avoid a wide application of compulsion. Even in a democratic régime a system of public control, instituted over the major economic activities of the citizens, must ramify until it embraces practically all the interests of a normal human life. To most men a society where the type of work in which a man may engage, the income which he receives, and the kinds of wants he is permitted to satisfy are all determined by authority bears little resemblance to freedom.

Planned economy and capital formation. The incentive of an interest return is believed to be necessary for the accumulation of capital in a system of free enterprise. The authoritarian experiments have shown that capital can be provided without appealing in this way to individual self-interest.

The soviet government in particular has certainly acquired a vast capital equipment. There has been some resort to borrowing

through interest-bearing government bonds, but the funds raised in this way have been only a small part of the total capital accumulation, and the purchase of the bonds by the people has been more largely a response to social pressure than to the attraction of the promised interest payment.

For the most part the capital has been provided by the people through savings which they either did not know they were making or which they could not avoid. Funds raised by taxation and goods acquired by governmental requisition are examples of compulsory savings. A more general form of saving, often unknown to the people who have borne the burden, has resulted from the government's power to determine the quantities of goods and services which were allowed to enter into direct consumption and the quantities to be employed in the production of capital instruments, a power which it possessed by reason of its control of the entire productive system.

To the average man in Russia the effect of this governmental policy has appeared in the form of a scarcity of consumption goods; he need not know that this scarcity merely reflected the government's decision to accumulate capital as rapidly as possible or, in other words, that he was bearing the burden of saving. It is in these ways that a communist régime, if given sufficient authority over the people, can meet its needs for capital equipment while dispensing with the incentive of personal award to the saver.

But in connection with this matter it should be noted that this system has no power to alter the process of capital formation in its essentials. The accumulation of capital involves a sacrifice of consumable income by some individual or group in society. Neither socialism, communism, nor fascism make saving burdenless. They can merely conceal the burden. This concealment is possible because the individuals who make the sacrifice do not take part in the decisions of the governing authorities. In Russia it is the dictatorial character of the soviet government which gives it power to accumulate capital without the knowledge or against the will of those who provide income for this purpose.

In a democratic system the people at large would be called upon to make the decision whether to live more comfortably in the present or provide for the increased comfort of later years and succeeding generations, and it is a safe conjecture that they would require some real incentive to induce them to forego present en-

joyment. The soviet leaders foretell the time when their governmental régime will be thoroughly democratic. Until the matter is put to a test at that time we have no evidence that democratic socialism—and this is the avowed objective of all socialist systems—will be able to acquire and maintain its indispensable capital equipment while abolishing pecuniary rewards to savers.

Planned economy and efficiency. Since comprehensive planning has been put to the test only in Russia, inferences regarding it must relate primarily to that country. It would appear that the soviet experiment in economic planning has been, on the whole, a success. It is true that things have not gone always according to prediction. There have been partial failures and unforeseen difficulties in many branches of the program, but this is to be expected of so gigantic and complex an undertaking. The planning commission has succeeded in guiding the nation's economic activity in general toward the goals set up in advance, and this achievement has been interpreted as evidence of a superior virtue in communism or socialism.

No such inference can be drawn from the facts. Systems of free enterprise proved equally capable during the war of mobilizing all their energies in support of national goals, and their reliance on individual initiative was far more effective in increasing production than the compulsions of authoritarian governments. This was but another example of the vitality of capitalism, its ability to adjust to altered conditions.

The point at issue between the two systems, however, is whether in normal times, not in times of emergency, more can be accomplished for the common economic good by leaving decisions to dictatorial public officials than to the people acting under the spur of self-interest. It is certainly a conservative appraisal of the record to say that the Russian experiment has not demonstrated the superiority of the former method. The success of the experiment in economic planning has been due to the dictatorship whose power stands behind it, and any equally strong dictatorship, whether of a socialist or a capitalist variety, would be able to accomplish similar results. Recalling again that communism in its true form presupposes a régime of democracy, the Russian experiment gives us no assurance as to the practicability of this aspect of the communist program.

Furthermore in Russia the attempt at economic planning has

been greatly facilitated by certain special conditions in the country. Throughout the entire period, the country has been laboring under conditions of extreme scarcity in all departments of its economic life. There has been a dearth of virtually every type of good and service, both for the satisfaction of human needs and for the development of basic productive equipment.

In many respects such a situation greatly simplifies the task of a planning commission. There is no problem of oversupply in any line of production; the authorities can safely assume a consumers' requirement larger than any product that it is possible for them to provide. There is no danger of an excess of labor in any branch of activity, since all are alike under-supplied with labor. No great thought need be given to the question whether the people would prefer more of this good and less of that, or would desire different styles or qualities of goods, since in the prevailing scarcity they will be happy to have any increase at all in the provision for their needs.

Obviously this condition is not typical of advanced industrial societies, where the problem is one of relative overproduction in certain lines of activity and relative oversupply of certain types of labor and capital, and where the preferential demands of the people must be considered. In these more typical situations, the task of a planning commission must be immensely more complex and difficult, and until the planned economy has been tested under these more typical conditions, it is unsafe to pass favorable judgment upon it.

What has been said applies in some measure also to economic planning in a fascist régime. It is unthinkable that anyone seriously desires the nations of the world to remain perpetually on a war footing, their entire life dominated by the voracious demands of militarism. What success economic planning had in Germany and Italy is pertinent almost exclusively to war. But the efficiency of a planned economy as compared with one of individual enterprise can be decided finally only in terms of the welfare of the people in peacetime. Neither communist nor fascist experiments give us evidence in support of the planned economy under normal conditions.

Suggestions for Further Readings

The books in the following bibliography are grouped under general headings, whose arrangement agrees with the order in which the respective subjects are treated in the text.

I. GENERAL ECONOMIC TEXTS

Slichter, S. H., *Modern Economic Society.* New York, 1931.
Ely, R. T., and others, *Outlines of Economics.* 6th ed. New York, 1937.
Garver, F. B., and A. H. Hansen, *Principles of Economics.* 3rd ed. Boston, 1947.
Meyers, A. L., *Elements of Modern Economics.* Rev. ed. New York, 1941.
Marshall, A., *Principles of Economics.* 8th ed. New York, 1920.
Taussig, F. W., *Principles of Economics.* 4th ed. 2 vol. New York, 1939.
Boulding, K. E., *Economic Analysis.* New York, 1941.
Bowman, M. J., and G. L. Bach, *Economic Analysis and Public Policy.* New York, 1943.
Smith, J. G., and others, *Economics and Social Institutions.* 5 vol. Boston, 1937–38.
Ise, J., *Economics.* New York, 1946.
Moffat, J. E., and others, *Economics, Principles and Problems.* 4th ed., New York, 1947.
Waugh, A. E., *Principles of Economics.* New York, 1947.
Tarshis, L., *The Elements of Economics.* New York, 1947.
Fairchild, F. R., E. S. Furniss and N. S. Buck, *Elementary Economics.* 5th ed. 2 vol. New York, 1948.

II. ECONOMIC JOURNALS

The American Economic Review. Menasha, Wis.
The Quarterly Journal of Economics. Cambridge, Mass.
The Journal of Political Economy. Chicago, Ill.
The Economic Journal. London, Eng.
The Annals of the American Academy of Political and Social Science. Philadelphia, Pa.
The Political Science Quarterly. Lancaster, Pa.
The Journal of the American Statistical Association. Menasha, Wis.
The Economist. London, Eng.
Econometrica. Menasha, Wis.
The Commercial and Financial Chronicle. New York, N. Y.
The Review of Economic Statistics. Cambridge, Mass.
The Federal Reserve Bulletin. Washington, D. C.
The Journal of Economic History. New York, N. Y.

III. FUNDAMENTAL CONCEPTS

Fraser, L. M., *Economic Thought and Language*, New York, 1937.
Cannan, E., *Review of Economic Theory*, London, 1929.
Robbins, L., *Nature and Significance of Economic Science*. New York, 1935.
Fowler, R. F., *The Depreciation of Capital*. London, 1934.
Fisher, I., *The Theory of Interest*. New York, 1930.
———, *The Nature of Capital and Income*, New York, 1906.
Commons, J. R., *Institutional Economics*. New York, 1934.
McConnell, D. W., and others, *Economic Behavior*. Rev. ed. Boston, 1939.
Edie, L. D., *Principles of the New Economics*. New York, 1922.
Dickinson, Z. C., *Economic Motives*. Cambridge, 1922.

IV. ECONOMIC HISTORY

Clough, S. B., and C. W. Cole, *Economic History of Europe*. Rev. ed. New York, 1946.
Ashley, W. J., *An Introduction to English Economic History and Theory*. New York, 1910.
Day, C., *Economic Development in Europe*. New York, 1942.
———, *History of Commerce*. 4th ed. New York, 1938.
Wright, C. W., *Economic History of the United States*. New York, 1941.
Williamson, H. F. (Editor), *Growth of the American Economy*. New York, 1944.
Kirkland, E. C., *History of American Economic Life*. Rev. ed. New York. 1939.
Shannon, F. A., *America's Economic Growth*. Rev. ed. New York, 1940.
Faulkner, H. U., *American Economic History*. 5th ed. New York, 1943.
Bogart, E. L., *Economic History of the United States*. Rev. ed. New York, 1938.

V. PRODUCTION

Black, J. D., *Introduction to Production Economics*. New York. 1926.
———, and A. G. Black, *Production Organization*. New York, 1929.
Snyder, C., *Capitalism the Creator*. New York, 1940.
Dewhurst, J. F., and others, *America's Needs and Resources*. New York, 1947
Nourse, E. G., and others, *America's Capacity to Produce*. Washington, 1934.
Leven, M., H. G. Moulton, and C. Warburton, *America's Capacity to Consume*. Washington, 1934.
Burns, A. F., *Production Trends in the United States Since 1870*. New York, 1934.
Moulton, H. G., *The Formation of Capital*. Washington, 1935.
———, and others, *Capital Expansion, Employment, and Economic Stability*. Washington, 1940.
Bell, S., *Productivity, Wages, and National Income*. Washington, 1940.
National Resources Committee, *The Structure of American Economy*, Part I. Washington, 1939.

Temporary National Economic Committee, 76th Congress, *The Structure of Industry*, Monograph No. 27. Washington, 1941.

———, *Technology in Our Economy*, Monograph No. 22. Washington, 1941.

VI. CORPORATIONS, ACCOUNTING, AND INVESTMENT

Purdy, H. L., M. L. Lindahl, and W. A. Carter, *Corporate Concentration and Public Policy*. New York, 1942.

Buchanan, N. S., *The Economics of Corporate Enterprise*. New York, 1940.

Seager, H., and C. A. Gulick, *Trust and Corporation Problems*. New York, 1929.

Hoagland, H. E., *Corporation Finance*. 3rd ed. New York, 1947.

Dewing, A. S., *Financial Policy of Corporations*. 4th ed. 2 vol. New York, 1941.

Haney, L. H., *Business Organization and Combination*. New York, 1934.

Gerstenberg, C. W., *Financial Organization and Management of Business*. 2nd rev. ed. New York, 1939.

Guthmann, H. G., and H. E. Dougall, *Corporate Financial Policy*. New York, 1940.

Kester, R. B., *Principles of Accounting*. 4th ed. New York, 1939.

Paton, W. A., *Essentials of Accounting*. New York, 1938.

Finney, H. A., *Principles of Accounting: Introductory*. New York, 1940.

Mills, F. C., *Statistical Methods*. New York, 1938.

Neiswanger, W. A., jr., *Elementary Statistical Methods*. New York, 1943.

Berle, A. A., and G. C. Means, *The Modern Corporation and Private Property*. New York, 1933.

Twentieth Century Fund, Inc., *Big Business, Its Growth and Its Place*. New York, 1937.

VII. THE FREE ECONOMY

Wriston, H. M., *Challenge to Freedom*. New York, 1943.

Harper, F. A., *The Crisis of the Free Market*. New York, 1945.

Lutz, H. L., *Guideposts to a Free Economy*. New York, 1945.

Wootton, B. A., *Freedom Under Planning*. Chapel Hill, 1945.

VIII. PRICE

Henderson, H. D., *Supply and Demand*. New York, 1922.

Robinson, E. A. G., *The Structure of Competitive Industry*. 6th ed. London, 1946.

Meade, J. E., *Introduction to Analysis and Policy*. 2nd ed. New York, 1947.

Stigler, G. J., *The Theory of Price*. New York, 1946.

Phelps-Brown, E. H., *The Framework of the Pricing System*. London, 1936.

Nourse, E. G., and H. B. Drury, *Industrial Price Policies and Economic Progress*. Washington, 1938.

Marshall, A., *Principles of Economics*. 8th ed. London, 1920.

Pigou, A. C., *The Economics of Welfare*. London, 1920.
Chamberlin, E. H., *The Theory of Monopolistic Competition*. 5th ed. Cambridge, 1946.
Robinson, J., *The Economics of Imperfect Competition*. New York, 1933.
Mills, F. C., *The Behavior of Prices*. New York, 1927.
Schultz, H., *The Theory and Measurement of Demand*. Chicago, 1938.
———, *Statistical Laws of Demand and Supply*. Chicago, 1928.
Burns, A. R., *The Decline of Competition*. New York, 1936.

IX. MONEY AND BANKING

Chandler, L. V., *An Introduction to Monetary Theory*. New York, 1940.
Bogen, J. I., and others, *Money and Banking*. 3rd ed. New York, 1947.
Kemmerer, E. W., *The ABC of the Federal Reserve System*. Princeton, 1938.
Thomas, R. G., *Our Modern Banking and Monetary System*. New York, 1942.
Prather, C. L., *Money and Banking*. 3rd ed. Chicago, 1946.
Westerfield, R. B., *Money, Credit and Banking*. Rev. ed. New York, 1947.
Halm, G. N., *Monetary Theory*. 2nd ed. Philadelphia, 1946.
Gayer, A. D., *Monetary Policy and Economic Stabilization*. 2nd ed. New York, 1937.
Angell, J. W., *The Behavior of Money*. New York, 1936.
Fisher, I., *The Purchasing Power of Money*. New York, 1920.
Hawtrey, R. G., *Gold Standard*. 5th ed. New York, 1947.
Reed, H. L., *Money, Currency and Banking*. New York, 1942.

X. PUBLIC FINANCE. GENERAL TEXTS

Schultz, W. J., *American Public Finance*, 3rd ed. rev. New York, 1942.
Lutz, H. L., *Public Finance*. 4th ed. New York, 1947.
Buehler, A. G., *Public Finance*. 2nd ed. New York, 1940.
Groves, H. M., *Viewpoints on Public Finance*. New York, 1947.
———, *Financing Government*. Rev. ed. New York, 1945.
Bullock, C. J., *Selected Readings in Public Finance*. 3rd ed. Boston, 1924.
Fagan, E. D., and C. W. Macy (eds.), *Public Finance*. New York, 1934.
Mills, M. C., and C. W. Starr, *Readings in Public Finance and Taxation*. New York, 1932.

XI. TAXATION

Green, W. R., *The Theory and Practice of Modern Taxation*. 2nd ed. rev. Chicago, 1938.
Stamp, J., *The Fundamental Principles of Taxation*. Rev. ed. New York, 1936.
Mering, O. O. von, *The Shifting and Incidence of Taxation*. Philadelphia, 1942.
Spaulding, H. B., *The Income Tax in Great Britain and the United States*. London, 1927.
Leland, S. E., *The Classified Property Tax in the United States*. Boston, 1928.

Jensen, J. P., *Property Taxation in the United States.* Chicago, 1931.
Twentieth Century Fund, Inc., *Facing the Tax Problem.* New York, 1937.
———, *Studies in Current Tax Problems.* New York, 1937.
Fairchild, F. R., and Associates, *Forest Taxation in the United States* (Report of the United States Forest Taxation Inquiry). Washington, 1935.

XII. GOVERNMENT EXPENDITURES, BORROWING, AND BUDGETS

Guest, H. W., *Public Expenditures.* New York, 1928
Stourm, R., *The Budget* (English translation). New York, 1917.
Committee on Postwar Tax Policy, *A Tax Program for a Solvent America.* New York: 1st Report, 1945; 2nd Report, 1947.
Buck, A. E., *The Budget in Governments of Today.* New York, 1934.
———, and others, *Municipal Finance.* New York, 1926.
Adams, H. C., *Public Debts.* New York, 1890.
Willoughby, W. F., *The National Budget System.* Baltimore, 1927.
Moulton, H. G., *The New Philosophy of Public Debt.* Washington, 1943.
———, and L. Pasvolsky, *War Debts and World Prosperity.* Washington, 1932.
Grady, H. F., *British War Finance, 1914–1919.* New York, 1927.
Jeze, G. P. A., and H. Trucky, *The War Finance of France.* New Haven, 1927.
Hansen, A. H., *Fiscal Policy and Business Cycles.* New York, 1941.
Harris, S. E., *The National Debt and the New Economics.* New York, 1947.

XIII. THE DISTRIBUTION OF WEALTH AND INCOME

Bowley, A. L., *The Division of the Profit of Industry.* Oxford, 1919.
Fellner, W., and B. F. Haley, *Readings in the Theory of Income Distribution.* Philadelphia, 1946.
Stigler, G. J., *Production and Distribution Theories.* New York, 1941.
Mills, F. C., *Economic Tendencies in the United States.* New York, 1932.
National Industrial Conference Board, *National Income in the United States, 1799–1938.* New York, 1939.
Kuznets, S., *Studies in Income and Wealth.* New York, 1937.
———, L. Epstein, and E. Jenks, *National Income and Its Composition, 1919–1938.* New York, 1941.
Leven, M., and K. R. Wright, *Income Structure of the United States.* Washington, 1938.
Doran, H. B., and A. G. Hinman, *Urban Land Economics.* New York, 1928.
Ely, R. T., and G. S. Wehrwein, *Land Economics.* New York, 1947.
Fisher, I., *The Theory of Interest.* New York, 1930.
Lyon, H., *Determinants of the Rate of Interest.* New York, 1942.
Meade, J. E., *Rate of Interest in a Progressive State.* London, 1933.
Moulton, H. G., *The Formation of Capital.* Washington, 1935.
Douglas, P. H., *The Theory of Wages.* New York, 1934.
Hicks, J. R., *The Theory of Wages.* London, 1932.

Knight, F. H., *Risk, Uncertainty, and Profit.* New York, 1921.
Fairchild, F. R., *Profits and the Ability to Pay Wages.* Irvington, N. Y., 1946.

XIV. INTERREGIONAL TRADE

Enke S., and V. Salera, *International Economics.* New York. 1947.
Ellsworth, P. T., *International Economics.* New York, 1938.
Haberler, G., *Theory of International Trade.* London, 1936.
Heilperin, M. A., *The Trade of Nations.* New York, 1947.
League of Nations, *The Network of World Trade.* New York, 1942.
Ohlin, B., *Interregional and International Trade.* Cambridge, 1933.
Hoover, C. B., *International Trade and Domestic Employment.* New York, 1945.
Young, J. P., *The International Economy.* New York, 1942.
Lary, H. B., *The United States in the World Economy.* Washington, 1943.

XV. FOREIGN EXCHANGE

Escher, F., *Modern Foreign Exchange.* New York, 1932.
Southard, F. A., P. F. Swart, and A. N. Gentes, *Foreign Exchange Practice and Policy.* New York, 1940.
Angell, J. W., *The Theory of International Prices.* Cambridge, 1926.
Buchanan, N. S., *International Investment and Domestic Welfare.* New York, 1945.
League of Nations, *International Currency Experience.* New York, 1944.
Heilperin, M. A., *International Monetary Reconstruction.* New York, 1945.
———, *International Monetary Economics.* New York, 1939.

XVI. COMMERCIAL POLICY AND PROTECTIVE TARIFF

Staley, E., *World Economic Development.* 2nd ed. Montreal, 1945.
Taussig, F. W., *Tariff History of the United States.* 8th ed. New York, 1931.
Beveridge, Sir W. H., *Tariffs, the Case Examined.* 2nd ed. New York, 1932.
Patten, S. N., *The Economic Basis of Protection.* Philadelphia, 1890.
Hirst, F. W., *Safeguarding and Protection in Great Britain and the United States.* New York, 1927.
Dixwell, G. B., *The Premises of Free Trade Examined.* Cambridge, Eng., 1881.
Gordon, M. S., *Barriers to World Trade.* New York, 1941.
League of Nations, *Commercial Policy in the Interwar Period.* New York, 1942.
Towle, L. W., *International Trade and Commercial Policy.* New York, 1947.
Isaacs, A., *International Trade and Commercial Policies.* Chicago, 1948.

XVII. BUSINESS CYCLES

Bratt, E. C., *Business Cycles and Forecasting.* Rev. ed. Chicago, 1940.
Clark, J. M., *Strategic Factors in Business Cycles.* New York, 1934.
Fisher, A. G. B., *Clash of Progress and Security.* New York, 1936.
Flanders, R. E., *Platform for America.* New York, 1936.

Mitchell, W. C., *Business Cycles, the Problem and Its Setting*. New York, 1927.
———, and A. F. Burns, *Measuring Business Cycles*. New York, 1947.
Robbins, L., *The Great Depression*. New York, 1934.
Simons, H. C., *Positive Program for Laissez Faire*. Chicago, 1934.
Swanson, E. W., and E. P. Schmidt, *Economic Stagnation or Progress*. New York, 1946.
Estey, J. A., *Business Cycles*. New York, 1941.

XVIII. LABOR PROBLEMS

Commons, J. R., and Associates, *History of Labor in the United States*. New York, 1932.
Daugherty, C. R., *Labor Problems in American Industry*. 5th ed. Boston, 1941.
Lester, R. A., *Economics of Labor*. New York, 1941.
Millis, H. A., and R. E. Montgomery, *The Economics of Labor*, 3 vol. New York, 1938 and 1946.
Peterson, F., *Survey of Labor Economics*. New York, 1947.
Bakke, E. W., *The Unemployed Worker*. New Haven, 1940.
Robinson, J., *Essays in the Theory of Employment*. 2nd ed. New York, 1947.
Pigou, A. C., *Employment and Equilibrium*. New York, 1941.
———, *The Theory of Unemployment*. New York, 1933.
Douglas, P. H., *Social Security in the United States*. 2nd ed. New York, 1939.
———, and A. Director, *The Problem of Unemployment*. New York, 1934.
Abbott, E., *Women in Industry*. New York, 1913.
Hobbs, C. W., *Workmen's Compensation Insurance*. 2nd ed. New York 1939.
Downey, E. H., *Workmen's Compensation*, New York, 1924.
New York State Department of Labor, *Hours of Work in Relation to Health and Efficiency*. Albany, 1941.

XIX. ORGANIZED LABOR MOVEMENT

Perlman, S., *History of Trade Unionism in the United States*. New York, 1937.
Peterson, F., *American Labor Unions*. New York, 1945.
Andrews, J. B., *Labor Problems and Labor Legislation*. 4th ed. New York, 1932.
Wolman, L., *Labor and the State*. New York, 1940.
———, *Ebb and Flow in Trade Unionism*. New York, 1936.
Commons, J. R., and J. B. Andrews, *Principles of Labor Legislation*. 4th rev. ed. New York, 1936.
Slichter, S. H., *Union Policies and Industrial Management*. Washington, 1941.
Witte, E. E., *The Government in Labor Disputes*. New York, 1932.
Taylor, A. G., *Labor Problems and Labor Law*. New York, 1938.
Bowman, D. O., *Public Control of Labor Relations*. New York, 1942.

Metz, H. W., *Labor Policy of the Federal Government.* Washington, 1945.
Twentieth Century Fund, Inc., *How Collective Bargaining Works.* New York, 1942.
Yellen, S., *American Labor Struggles.* New York, 1936.

XX. AGRICULTURE

Taylor, H. C., *Outlines of Agricultural Economics.* New York, 1931.
National Industrial Conference Board, *American Agricultural Conditions and Remedies.* New York, 1936.
Schultz, T. W., *Agriculture in an Unstable Economy.* New York, 1946.
Clark, F. E., and L. D. H. Weld, *Marketing Agricultural Products in the United States.* New York, 1932.
Nourse, E. G., J. A. Davis and J. D. Black, *Three Years of Agricultural Adjustment Administration.* Washington, 1937.
Black, J. D., *Parity, Parity, Parity.* Cambridge, 1942.
———, *The Dairy Industry and the AAA.* Washington, 1935.
Richards, H. I., *Cotton under the Agricultural Adjustment Act: Developments up to July, 1934.* Washington, 1934.
Davis, J. S., *Wheat and the AAA.* Washington, 1935.
Rowe, H. B., *Tobacco under the AAA.* Washington, 1935.
U. S. Department of Agriculture, "Farmers in a Changing World." *1940 Yearbook of Agriculture.* Washington, 1940.

XXI. TRANSPORTATION

Daggett, S., *Principles of Inland Transportation.* 3rd ed. New York, 1941.
Frederick, J. H., *Commercial Air Transportation.* Rev. ed. Chicago, 1946.
Healy, K. T., *Economics of Transportation in America.* New York, 1940.
Jones, E., *Principles of Railway Transportation.* New York, 1924.
Bigham, T. C., *Transportation: Principles and Problems.* New York, 1946.

XXII. INDUSTRY AND TRADE

American Economic Association, *Readings in the Social Control of Industry.* Philadelphia, 1942.
Purdy, H. L., M. L. Lindahl, and W. A. Carter, *Corporate Concentration and Public Policy.* New York, 1942.
Ripley, W. Z. (editor), *Trusts, Pools, and Corporations.* Boston, 1905.
Curtis, R. E., *The Trusts and Economic Control.* New York, 1931.
Jones, E., *The Trust Problem in the United States.* New York, 1921.
Laidler, H. W., *Concentration of Control in American Industry.* New York, 1931.
Arnold, T. W., *The Bottlenecks of Business,* New York, 1940.
Burns, A. R., *The Decline of Competition.* New York, 1936.
Seager, H. R., and C. A. Gulick, *Trusts and Corporation Problems.* New York, 1929.
Jenks, J. W., and W. E. Clark, *The Trust Problem.* New York, 1929.

Lyon, L. S., and others, *The National Recovery Administration: An Analysis and an Appraisal.* Washington, 1935.
Grether, E. T., *Price Control under Fair Trade Legislation.* New York, 1939.
Stevens, W. H. S., *Unfair Competition.* Chicago 1917.
Miller, J. P., *Unfair Competition.* Cambridge, 1941.
Bonbright, J. C., and G. C. Means, *The Holding Company.* New York, 1932.
Clark, J. M., *Social Control of Business,* 2nd ed. rev. New York, 1939.

XXIII. CONTRASTING ECONOMIC SYSTEMS

Marx, K. (edited by A. Lee), *Essentials of Marx.* Rev. 2nd ed. New York, 1946.
Lenin, N., *Imperialism. The State and Revolution.* New York, 1926.
Laski, H. J., *Communism.* New York, 1927.
Pigou, A. C., *Socialism versus Capitalism.* New York, 1937.
Cole, G. D. H., *Guild Socialism.* London, 1920.
Hoover, C. B., *Economic Life of Soviet Russia.* New York, 1931.
Webb, B., and S. Webb, *Soviet Communism.* 3rd ed. New York, 1945.
———, *A Constitution for the Socialist Commonwealth of Great Britain.* London, 1920.
Loucks, W. N., and J. W. Hoot, *Comparative Economic Systems.* Rev. ed. New York, 1948.
Westmeyer, R. E., *Modern Economic and Social Systems.* New York, 1940.
Bavkov, A. M., *Development of the Soviet Economic System.* New York, 1946.
Gibberd, K., *Soviet Russia.* Rev. ed. London, 1946.
Schuman, F. L., *Soviet Politics, at Home and Abroad.* New York, 1946.
Vernadsky, G., *History of Russia.* Rev. ed. New Haven, 1944.

Index

Ability to pay, 486–91
Acceptance, 287–88
Accounting, for corporation, 45–54; for individual proprietorship, 54; for partnership, 54
Accidents, industrial, 678–81
Advantage, law of comparative, 532–39
Advertising, 207–208
Aged, pensions for, 674–75
Agricultural Adjustment Act of 1933, 728–32; objectives of, 728; control of production under, 729; and agricultural recovery, 731; and business recovery, 731–32; comparison with act of 1938, 733
Agricultural Adjustment Act of 1938, 733–36; comparison with act of 1933, 733; "ever-normal granary" of, 734–35
Agricultural Marketing Act of 1929, 726–27
Agriculture, and large-scale production, 61, 161–62; increasing costs in, 176–77; government aid to, 343–44, 730–33; prices in, 396, 721–23, 725–31, 738–42; rent in, 394–401; problems of American, 719–43; economic peculiarities of, 719–21; during the First World War, 721–23; incomes in, 721–23, 734, 737–38, 741–42; in 1920's, 723–25; depression in, 725; federal programs for, 726–40; debt in, 726, 738; relief, 1921 to 1929, 726-27; marketing act of 1929 for, 726–27; control of production in, 729, 731, 740; acreage allotments in, 729–31, 734; insurance in, 734–36; disposition of surplus commodities in, 736; expansion of, 738; appraisal of government policy in, 740–43; production during Second World War, 740; in Russia, 821
Agricultural credit agencies, 343–44
Air carriers, 768
American Federation of Labor, 684–86; political activity of, 692
Accidents, compensation for industrial, 678–81
Anti-inflation act of 1942, 739–40
Anti-trust laws, 770–71, 772–79, 788–89
Apportionment of property tax, 462–63
Arbitrage, 539
Arbitration, 709–18; voluntary, 710–12; compulsory, 714–18; appraisal of compulsory, 717
Assessment, of property tax, defined, 462; organization, procedures, and results of, for property tax, 463–66; special, 457–58
Assets, defined, 45
Auction, common, 93–94; buyer's, 94–96

Balance of payments, international, 569–73; of the United States, 570–71
Balance sheet, explained, 45–49; bank's, 288–92
Bank credit, 282–312; importance of, 282–83; elasticity of, 308–10; qualitative or quantitative control of, 327–30; expansion of, before 1929, 335–36; liquidation of, after 1929, 336–37; government support of, 342–43; for agriculture, 343–44; for real estate, 349; for speculation, 349–50
Banking act of June, 1933, 271, 273
Banking act of March 9, 1933, 270
Banking, principles of, 280–312; branch, 333–34, 338–39; chain, 334;

American system of, 312–31; history of American, 313–18; recent developments in American, 332–52; weakness of American, 334–35, 351–52; legislation concerning, since 1929, 338–41; functions assumed by American government, 341–52; criticism of government, 345–46
Bank notes, national, 315–16, 324–26; compared with deposits, 306–308; defined, 306; regulation of, 307–308; federal reserve, 324–26; elasticity of, 324
Bank of the United States, first, 313–14; second, 314
Bank reserves. *See* Reserves
Banks, central, 313–15
Banks, closing of (March 1933), 270; types of, 283–84; reserve ratios of, 300–305; American local (state), 313–15, 332–33; mutual savings, 332–33; number of, 332–33; private, 332–33; failures of, 337–38, 341–42; capital requirements of national, 339–40; and RFC, 341–43; reopening of (1933), 342; for coöperatives, 544; trends in operations of commercial, 352; and foreign exchange, 559–60
Banks, home loan, 344–45
Banks, land, 343
Barter, 14, 253–54
Bill of exchange, 556, 558; determination of rate by demand and supply of, 560–61
Bill of lading, 556
Bimetallism, 263–67; adjustment of discrepancy in ratios under, 265–67; Thomas amendment permitting, 274
Bonds, explained, 43–44; payment of interest on, 43, 49; and stock compared, 42; government, 790–91
Borrowing, spendthrift, 383; for consumption and time preference, 385–86; purposes of, 385–86; by government, 386–87, 521–25; for agricultural production, 733
Bounties, protection by, 610–11
Brassage, 275n
Bretton Woods Agreements, 553n, 590, 593–94
Brookings Institution, 730
Budget, explained, 525–26; of United States government, 525–26; of Great Britain, 526; effect of Legislative Reorganization Act on U. S., 527; balancing of U. S. government, 527–30
Business activity, index of, 617
Business cycle, 614–39; and accidental irregularities, 615; characteristic features of, 616–19; revival, 616, 621–22; recession, 617, 624–25; depression, 617, 626; major and minor, 618–19; variations in, 619–20; characteristic casual relationship of, 620; prosperity, 622–24; forces initiating, 627; theory of matured economy, 628–31; restrictions on free economy, 631–33; and unemployment, 640–42, 646; socialist explanation of, 804–805
Business firm, function of, 35–36
Business management. *See* Entrepreneur
Business taxes, 478–79
By-products, and large-scale production, 64

Canada, compulsory investigation of labor disputes in, 713
Canal, development of early, 745–46
Capital, a factor of production, 21–23; accumulation of, 23–25; defined, 47n, 231n; growth of, 57–59; and income, 231–32, 238; and property, 232–33; valuation of, 233–50; income of, 374n; accumulation of, 375; non-business demands for, 383–87; time preference and formation of, 410–11; formation during 1930's, 629, 631–35; accumulation of, in Russia, 824–25, 829–30
Capitalism, socialist view of, 798–801 *See also* Free Economy
Capitalization, of taxes, 481 *See also* Discount
Capital stock, 41–43
Chain store, 785; taxation of, 785–86; legislation directed at, 785–86
Check, bank, 293–96
Chicago Board of Trade, 96, 174–75
Child labor, 663–67; use of taxation to regulate, 505; characteristics of, 663–64; extent of 664–65; effects of, 665;

legal regulation of, 666–67; unconstitutionality of laws regulating, 667; bargaining power of, 670–71; wages of, 670
"Class rates," railroad, 759
Class struggle, 801–802
Clayton act of 1914, 770, 774, 777–78
Clearing house, 296–99
Clearings, bank, 295–99; by Federal Reserve, 352
Closed shop, 698–700
Closed union, 698
Coal industry, regulation of, 783–84
Codes, of fair competition, 782–83; labor, 782–83
Coinage, 261; free, 261; gratuitous, 261; under bimetallism, 263–65; of silver dollars, 275, 277–78
Collateral security, 556
Collective bargaining, 694–98; of child and women labor, 670–71; legality of, 687–88; under Labor Management Relations Act, 690; benefits to employers of, 695–96; conflict over method in, 696–97
Colorado, Industrial Commissions Act, 713
Combination, and large-scale production, 56–74; defined, 67; formation of, 67–68; horizontal, defined, 68; motives for, 68; economies of, 69–71; vertical, defined, 71; technical advantages of vertical, 71; weaknesses of specialization in, 72–74; earnings of, 72–73
Commercial banks, operations of, 283–300; balance sheet of, 288–92; collection between, 295–99; volume of clearings of, 298; increasing importance of investments of, 299–300; portfolio of, 299–300; number and size of, 332–33; legislation concerning, since 1929, 338–41; divorced from investment banking, 340–41; trends in operations of, 352
Committee on Industrial Organization, 685–87; political activity of, 692–93
Commodity Credit Corporation, 733, 742
"Commodity dollar," 370
"Commodity rates," railroad, 759–60
"Commodity reserve" scheme, 371
Common law, 38n; in partnerships, 38; of negligence in industrial accidents, 679–80
Common stock, 42
Consumer protection of, 779–80
Communism, 812–26; hostility to free enterprise, 81; doctrines of, 813–14; organization in Russia, 814–16; international organization of, 816–18; program of, 818–26; and fascism, 826–27
Communist unions, 823
Conciliation, in labor disputes, 706–708
Cooperation, 27; simple, 27–28; by division of labor, 28–29
Company unions, 690–91
Comparative advantage, principle of, 532–39, 595, 597
Compensation laws, workmen's, 678–81; constitutionality of, 680–81
Competing cost goods, prices of, 244–45
Competing goods, prices of, 243–44
Competition, 76–77, 100; types of, 100; supply under, 168–84; pure and imperfect compared, 200, 211; quality and price in, 202–203; "cutthroat," 220–21; benefits of foreign, 605; in transportation, 752–58, 762; maintenance of, 769–72; and monopoly, 769, 772; restraint of, 770–72; "fair," 771; and mergers, 773; prevention of unfair, 778; plane of, 779–80; legislation to restrain, 780–88; unfair methods of, 780; under NIRA, 782–83; in soft coal industry, 783–84; socialist criticism of, 797–99, 806
See also Imperfect competition; Pure competition
Competitive price, and public interest, 87; laws of, 111–12; theory and practice, 246–50
Complementary goods, prices of, 244
Compulsory investigation, of labor disputes, 712–14
Constant costs, conditions essential for, 177
Constitution of United States, and income tax, 468–69; limitations of taxation, 482–83; and taxation, 506–507; proposed child labor amend-

ment to, 667; and workmen's compensation laws, 680–81; and National Labor Relations Act, 689; processing taxes, 732; and NIRA, 783
Construction, during 1930's, 636–37
Consumer, influence of, on price, 85, 90, choices, 125
Corporation, importance of, 38–39; nature of, 39–41; forms of securities of, 41–44; property account of, 45–49; income account of, 51–54; power of, to own stock, 67–68; valuation of the, 248
See also Corporations; Trusts; Holding company
Corporations, assets of largest, 65–67; growth of large, 66–67; concentration of, control of, 66–67; income of, 66; savings by, 412–13; taxation of, 466–72, 494; abuses by, 792
Cost, defined, 9; relationship to supply, 156–57; effect of future, on supply, 157–58; fixed, 158–59, variable, 158–59, 171; effect of expansion of output on, 159–61; when plant is fully utilized, 159–60; analysis of, 159; average, 162–64, 168–70; marginal, 162–66; relation to output, 164–66; schedules and curves, 164–65; effect upon seller's policy when price above average, 168–70; effect upon seller's policy when price below average, 170–71; long-run and short-run, 171–72; effect on volume of production, 171–72; increasing, 175–77; constant, 177–80; decreasing, 182–84; of production and discounted income, 240–46
See also Marginal cost
Cost of goods sold, 52
Costs, general and administrative, 52; selling, 52; other, 52; of government, 454–55; in agriculture, 722–23; of transportation, 749–52
Cotton, government measures affecting, 728–33
Credit, 280–83; commercial, 281–82; function of, 282–83; depends on savings, 282–83; exchange through bank discount and deposit, 292–93; elasticity of bank, 308–10; control of, 327–30; control of rediscount rates, 328; control by open-market operations, 328–30; control by regulation of reserve ratios, 330; for agriculture, 343–44; for housing, 344–45; speculative use of, 349–50; effect on gold movements of transfers of, 567
See also Bank credit
Credit money, 258; irredeemable, 267–69
Crop insurance, 734–36
Currency, based on money, 255–56; defined, 256; need of elastic, 308; quantity of, and value of money, 361–63; velocity of circulation of, and value of money, 363
Customs tariff, 473–77; income from, 476–77; shifting and incidence of, 480; use by federal government, 515–17
Cycle, *see* Business cycle

Death taxes, 477–78
Debt, war, 337; and fluctuating value of money, 367; moratoria of international, 585–86; in agriculture, 722, 726, 738
See also Government debt
Decreasing-cost industries, 182–84
Deed, 13–14
Deficit, government, 455
Demand, 125–151; and large-scale production, 61; and price, 102–103; schedule of, 103–104, 139–43; curve of, 105; and supply together, 107–11; laws of supply and, 111–12; two possible meanings of, 112–14; effect of changes on price and quantity exchanged, 115–19; relation to marginal utility, 133–35; and desire, 135–36; relation of marginal utility to, 136–39; individual, 136–39; rational basis of, 138–39; defined, 139–40; total, 139; actual schedules, 141–43; law of, 143; elasticity of, 146–51; and price under imperfect competition, 192–94; importance of ascertaining facts of, 203–205; stimulation of, 205–208; under imperfect competition with few sellers, 217; under monopoly, 227–28; partly determined by discount, 238–40; for productive

factors, 374–92, laws of, for productive factors, 381, 389–90
Deposit insurance, 348–49; appraisal of, 348–49; and government regulation of banks, 352
Deposits, a form of currency, 255–56; pooling of savings by, 282–83; defined, 292; demand and time, 292; comparison with bank notes, 306–308; elasticity of, 309–10; reserves against, 326–27; government regulation of interest on, 346; and value of money, 361–63; velocity of circulation of, 363
Depreciation, of dollar, 271–73; and interest, 418
Depression, of 1929, 268–70, 336–37; and federal expenditure, 511; and federal revenues, 516–17; in business cycle, 617, 621, 626; of the 1930's, 628–29; in agriculture, 723–25; and destructive competition, 780–81
Devaluation, of the dollar, 271–73
Dictatorship, hostility to free enterprise, 81
Diminishing returns, 379–80; law of, 177, 380; importance of principle of, 380; and borrowing, 385–86; in agriculture, 394–96; and marginal cost, 395–96; on urban lands, 402–403
See also Marginal cost
Discount, formula, 235–37; process of, 237–38; and laws of price, 238–40; creation of credit by bank, 283; bank, 285–86; by federal reserve banks, 321–22; of speculative loans, 349–50
Discrimination, in rail rates, 755–58
Distribution, meaning of, 374; functional, 374; and process of production, 376–77; of income, 390–92; rise of mass, 784–85
See also Taxation
Dividends, declaration of, 42; payment of, 49–50; stock, 49–50; taxation of corporation, 470–71
Division of labor, 28–33; within trades, 28; by trades, 28; economies of, 29–30; effects on labor, 30–31; and extent of market, 31; territorial, 31–32; a form of coöperation, 32–33; and interregional trade, 531–32; extension by multilateral trade, 537–39; international, 597
Dollar, devaluation of, 271–73; stabilization of exchange value of, 273–74; commodity or stabilized, 368–73
Draft, 287–88; demand or sight, 558
"Dumping," and protective tariff, 607, 608

Economic fluctuations, 614–16
Economic freedom, and the control of production, 75–87
Economic nationalism, *See* Nationalism, economic
Economic rent, 393–408
See also Rent
Economics, foundations of the science of, 1–16; defined, 4–5; a science, 5–6
Elasticity, of demand, 146–49; of supply, 181–82; of supply, influences affecting, 181; of supply, limits of, 182
Employment, index of factory, 618; during the 1930's, 628–29; fluctuations in 640–42; part-time, 642; seasonal fluctuations in, 645–46
See also Unemployment
Employment offices, public, 649–50
Entrepreneur, 26, 375–76; function of, 26–27; and risk, 26, 444–45; a "factor of production," 27; function in business firm, 35; incentive of, 445–46
Entrepreneur-demand, law of, 380–81; total, 381
Equalization, of property tax, 462
Equities, defined, 45
Erdman act, 709
Estate tax, 477; federal, 516–17
Excise taxes, shifting and incidence of, 421; federal, 472–73, 515–17; state, 473
Exchange, 14, 96–97; determination of price on organized, 102–12
See also Foreign exchange; Bill of exchange; Trade; Exchange rate
Exchange Equalization Account of Great Britain, 369, 589
Exchange rate, 557; stabilization of, 369
See also Foreign exchange rate

Expenditures, government. *See* Government expenditures
Expense, general and administrative, 52; selling, 52; other, 53
Exports, importance of, 536–37; balancing imports and, 536–37, 549; world, 542; United States, 546–48; of U. S. in relation to production, 547–48; financing, 555–57; reaction of exchange rates on, 567–79; and antitrust laws, 778–79

Factors of production, business demand for, 378–81
Factory laws, relation to free enterprise, 79–80
Fair Labor Standards Act of 1938, 667, 672–73
Fair trade laws, state, 787–88
Farm. *See* Agriculture
Farm Credit Administration, 343, 727
Farm problem. *See* Agriculture
Farm Security Administration, 737
Fascism, 826–28; and communism, 826–28
Federal Coal Commission, 784
Federal Crop Insurance Corporation, 734
Federal Deposit Insurance Corporation, 346–48
Federal Farm Board, 726–27
Federal Home Loan Bank Administration, 345
Federal Housing Administration, 345
Federal Mediation and Conciliation Service, 709
Federal National Mortgage Association, 345
Federal Public Housing Authority, 345
Federal reserve banks, 318–31; discounting by, 321–22; rediscounting by, 321–22; notes of, 322–24; deposits in, 326; reserves of, 326–27; rediscount rates, 328; open-market operations of, 328–30; legislation concerning, since 1929, 338–41
Federal Reserve System, 318–31; establishment of, 318; organization of, 318–19; membership in, 319–21, 351; open market operations, 328; regulation of reserve ratios, 330; clearing system, 330–31; government fiscal agency, 331; powers of board of, 350, 352
Federal Trade Commission, 774; activities against restraint of trade, 777–78; and unfair methods of competition, 780
Federal Trade Commission act, 770–72
Fees, defined, 456–57; distinguished from taxes, 457–58; and special assessment, 457–58
Firm, the business, 35–54
Fluctuations, economic, 614–16
Food and Drug act of 1906, 779–80
Food and Drug act of 1938, 779–80
Foreign exchange, 555–69; governmental control of, 369–70; for transactions other than trade, 558; market in, 560; role of banks in, 559–60; sterling, 561, 588–89; indirect, 564–65; equilibrium prior to First World War, 575–76; war and post-war conditions of, 575–76; under irredeemable money standards, 576; purchasing power par of, 577–79; stabilization of, 583–88; government control of, 585–88; clearing system, 587–88; functions of International Monetary Fund in, 592–94; restrictions under International Monetary Fund, 593–94
See also, Foreign exchange rate; Exchange rate
Foreign exchange rate, defined, 560; explained, 560n; mint par of, 560–61; determination by supply and demand of bills, 560–61; gold points of, 561–64; limits to flow of gold, 565; effects of imports and exports on, 567–69; under purchasing power par, 578–79, 597–99
Foreign investments, 549–54; reasons for, 549–50; of the United States, 552–53
Foreign trade. *See* Interregional trade
France, government regulation in, 76
Freedom, economic, and control of production, 76–84
Free economy, 78; reform to make work, 79–82; reform to displace, 80–89; geographical extent, 82–83; recent restriction on, 631–33; eco-

nomic progress and the preservation of, 638–39; present status of, 796–97; adaptability of, 810–11
Free trade, 76–77, 595, 605–606; meaning of, 597n
Functional distribution, 374

General property tax, evolution of, 462; general description of, 462–66; assessment of, 462; importance in American tax systems, 463; organization, procedures, and results of assessment, 463–66; revenue from, 463; taxation of intangible property, 466; application to corporations, 466–67; reform of, 467–68
Ghent system, of unemployment relief, 656
Gold, why used as money, 259–60; as standard money, 260–67; and silver under bimetallism, 263–64; embargoes on exports of, 269; surrender of, in 1933, 270–71; hoarding of, 270–71, 276; clause concerning payment in, 273; policy of purchasing, 272; nationalization of, 272–73; Treasury's profit in, through devaluation, 273–74; bullion used to settle international balances, 276, 562–64; holdings of, of Federal reserve banks, 277–78; stock of, in United States, 277–78; international flow of, 562–64; effect on price levels of movements of, 566; "sterilizing," 566n; effects of credit transfers on movements of, 567; desirability as import, 572–73; world price of, 593
Gold bullion standard, 262
Gold certificates, 277–78
Gold exchange standard, 262–63, 582
Gold points, of foreign exchange, 561–64
Gold reserve act of January 30, 1934, 272–73, 589
Gold standard, 260–62; characteristics of, 260–61; Great Britain and, 260–61, 580–82; return to, during recovery, 269, 580; war-time desertion of, 269; abandonment of, following 1929, 269–70, 583; and foreign exchange rates, 561–63; and gold movements, 572–73
Government, functions of, under *laissez faire*, 77; control of money, 254–59, banking functions assumed by American, 315–31, 341–52; independent treasury system of U. S., 314, 331; Federal Reserve System the fiscal agent of U. S., 331; borrowing of, 386–87; its place in the economy, 452–53; grades of American, 453; organization of American, 453–54, American units included, 453n; functions of, 454; sources of income of, 454–55; costs defined, 455; financing primary functions of, 455; cost of, 502–14; borrowing by, 521–25; political activity of unions and structure of, 692
Government bonds, 790–91
Government debt, 521–25; security of, 522–24; of United States, 524–25
Government banking, criticism of, 345
Government budget. *See* Budget
Government expenditures, 508–14; increase of, 508–509; federal government, 509–14, 527–30; American state and local, 512–14; functional analysis of American, 512–14; distribution among grades, 512; during the 1930's, 628–29
Government industry, 789–95; financing of, 455; savings of interest and profits, 790–91; financial efficiency of, 792–93; motives in, 793–94; political interference in, 793; versus private industry, 795
Government regulation, of industry, 75; in Soviet Russia, 75; in Great Britain and France, 76; of prices, 89; of monopoly, 228, 268–75, 770–71, 772–79, 795; of note issue, 308, 309–10, 315–16; of banking, 338–49; of foreign exchange, 368–73; of child labor, 505, 666–67; of coal industry, 505–506, 783–84; of business cycle, 506; of trade, 609, 772–79; of collective bargaining, 697–98; of wages, 672–73, 782–83; of strikes, 706–18; of agricultural production, 729–31, 740; of transportation, 753–54, 762–66; of industry and trade, 769–95; types of, 769–72; reasons for, 769–72; of trusts, 770–71, 772–79; of securi-

ties, 771; of holding companies, 771–72; of competition, 783–88; of hours of labor, 782–83; of labor, 782–83
Government revenues, federal, 514–17, 527–30; state, 517–20; local, 520–21
Great Britain, government regulation in, 76; status in free economy, 82–83; preferential tariff arrangements, 611; loan and financial agreement of 1946, 613
"Gresham's law," 265
Guffey-Vinson act, 783–84

Higgling, 92–93
Highway system, topographic limitation of, 748–49
Holding company, 772; use by banks, 339; government regulation of, 771–72
Home loan banks, 344–45
Hours of work, 661–63; under Fair Labor Standards Act, 672–73
Housing, federal credit agencies for, 344–45

Imperfect competition, 101–102, 186–208; defined, 102, 190–91; conditions of, 186–92; individuality of product, 187–89; few sellers, 189–90, 215–23; large offering of one seller, 190; seller's problem under, 192–94; price and demand under, 192–94; marginal revenue under, 194–95; marginal cost and marginal revenue under, 195–97; maximum profits under, 197–99; seller's liberty under, 200–202; effect of buyer's preferences, 212; many sellers and differentiated product, 211–15; fewness of buyers, 223–24
See also Competition, Pure competition
Imports, importance of, 536–37; balancing exports and, 536–37, 549; U. S., 545–46; financing, 557–59; reaction of exchange rates on, 567–68
Incidence, of taxation, 479–81
Income, defined, 9; net, 9–10, 469; measurement of, 10, 10n; distinguished from wealth, 10; of corporations, 51–54; present value of future, 231–38, 383–86; and capital, 231–32, 238; types of, 374–76; implicit and explicit, 376; functional distribution and national, 390–92; surplus, 407–408; labor's share of, 425–40; law of proportionality and, 438–40; profits a residual, 443–44; sources of government, 454–55; gross, 469; national, 508, 725; in agriculture, 722–23, 725, 738, 741–42; and parity prices, 734, 741–42
See also Distribution
Income tax, historical background of, 468–69; and constitution of United States, 468–69; general description of, 469–72; individual, 469–72; exemptions, 469; corporation, 469–72; rates in United States, 469, 471; partnerships, 469; rates in Great Britain, 471; elasticity of, 471–72; revenue from, 472; shifting and incidence of, 479–81; federal, 516–17
Increasing costs, 175–77
Independent treasury system. *See* Government
Index numbers, 355–59
Individual proprietorship, 36
Industrial conflict, devices to prevent or settle, 706–18
Industrial unions. *See* Labor unions
Industry, geographical shifting of, 500
Inelasticity, of demand, 149–51
Inflation, 372, 576; effect on exchange rates, 578–79
Inheritance tax, 477–78
Injunction, in labor disputes, 706
Insurance, of bank deposits, 346–49; and time preference, 384, 384n; and interest, 418; of unemployment, 655–60; of unemployment in Great Britain, 656–58; of old age, 675–77; of industrial accidents, 680–81; in agriculture, 734–36
Intangible personal property, taxation of, 466–67
Interest, 410–24; defined, 234–35, 375; link between present and future values, 235–37; formula, 235; and discount, 285–86; unearned, 285; government regulation of rates on bank deposits, 346; rate of, 377–78; rate of, and government borrowing, 386; and loan market, 387, 411; im-

plicit and explicit, 392, 442–43; supply of loans and rate of, 413–14; quantity of money and rate of, 414–16; bank reserves and rate of 415–16; lack of uniformity in rates of, 416–17; true and nominal rates of, 417–18; and depreciation, 418; and wages of management, 418; and insurance, 418; as a private income, 420; regulation of, 421–22; interference with rates of, 421–22; savings of, 790–91
Interlocking directorates, 777–78
Intermediate goods, 21
International Bank for Reconstruction and Development, 553–54, 590
International Conference on Trade and Employment of 1947, 611–12
International Monetary Fund, 590–94; objectives, 590–91; organization, 591; operation, 591–92
International trade. *See* Interregional trade
Interregional investment, 549–54
Interregional trade, 531–54, 586–87; bases of, 531–32; and division of labor, 531–32; and production, 531–32; and industrialization, 540–41
Interstate commerce, regulation of, 185–86, 753–54, 762–66; minimum wages in industries in, 672–73
Interstate Commerce Commission, 753–54, 763; and rules for control of railroad rates, 763–64
Invention, 25–26, 643; a prerequisite of large-scale production, 60–61; and decreasing cost industries, 183–84
Investment, under-, 423; interregional, 549–54; foreign, 549–54; and revival, 621, 627; during recession, 624–25; alleged decline in opportunities for, 630; failure during 1930's of revival of, 631–37
Investment banking, divorced from commercial banking, 340–41
Irredeemable money, 257, 267–69

Joint cost goods, prices of, 245
Justice, in taxation, 492

Kansas, compulsory arbitration in, 715–16

Labor, 20–21, 375; non-business demand for, 382–83; demand for, 425–26, 642–43; the wage rates of, 425–40; supply of, 426–27, 642–43; wages and interdependence of types of, 428–29, its share of social income, 438–40; protection and demand for, 600; restrictive practices of, 636; reserve of, 646; immobility of, 648–49; stabilizing demand for, 651–52; hours of, 661–63; unemployment and defensive policies of, 646–48; in Russia, 822–24
See also Labor unions; Unemployment; Wages
Labor, child. *See* Child labor
Labor exchanges, 649–50
Labor Management Relations Act, 689–90; effect on company unions, 691; provisions concerning political activity, 692; union welfare funds, 694; collective bargaining provisions of, 697–98; closed shop provisions of, 700; strikes and lockouts, 706; conciliation and mediation provisions of, 709–710; compulsory investigation of labor disputes, 713–14
Labor movement, The American Organized, 682–718
Labor problems, other than unemployment, 661–81
Labor unions, unemployment insurance of, 655–56; conditioning factors, 682–83; defined, 683; company unions not included under, 683, craft or trade, 683–84; federation of, 684; industrial, 684–85; conflict of, rival types of, 685–87; legal position of American, 687–88; membership of, 687; legal responsibility of, 690; conflict with company unions, 691, political activity of 691–93; benefit systems of, 693–94; and collective bargaining, 694–98; recognition of, 694–95; welfare funds, 694; and wage rates, 694–95; closed, 698; and closed shop, 698–700; restrictive practices of, 700–701; and strike, 701–706
See also Company unions
Labor, woman. *See* Woman labor

Labour Party, of Great Britain, 808–809
Laissez faire, explained, 76–77; reaction from, 78–83
See also Free economy
Land, 20–21, 374–75; non business demands for, 387–89; immobility of, 389; productivity of, 393–94, 398–99; alternative uses of agrarian, 400–401; varying productivity of urban, 401–402; inequality of urban, 401–402; diminishing returns on urban, 402–403; value of, based on rent, 403–404; shifting of tax on value of, 480–81; taxation and utilization of, 501–502
Land banks, federal, 343
Large-scale production, and combination, 56–74; and the standard of living, 56–77; growth of, 59–60; importance of, 59–60; conditions necessary for, 60–61; prevalance among industries, 61, economies of, 61–64; limit to efficiency of, 161–62
Legal tender, 254–55; American, since 1933, 273
Legislative Reorganization Act of 1946, 527
Liability, unlimited, 36, 37; double, 40; limited, 40
Liabilities, defined, 45; types of, 45
License taxes, 478
Loans, security for, 287; payment of, 290–91; on real estate by banks, 349; and capital, 375; demand for, 378–87 consumption, 383–86; and rate of interest, 385, 387, 411; government, 386–87; supply of, 411–24; and bank reserves, 414–15; quantity of money and demand for, 415–16; interest on consumption, 421–22; "without recourse," 731; agricultural, 731–33
Lockout, 701
Long run, seller's policy in, 172; supply in, 180–81; equilibrium under imperfect competition, 213–15
Losses, 47–48; and profits of corporations, 446–49

Machinery, and large-scale production, 63
Maintenance of membership, clause in union contracts, 700
Management, importance in large-scale production, 61–63; and small-scale production, 65
See also Entrepreneur
Marginal cost, 162–66, 168–70, 395–96, 398; curve of, 165, 168–70; related to supply under constant costs, 178–79
Marginal productivity, of labor, 427–29; a limit to wage rates, 429–432; importance in fixing wages, 436–37; defined, 167; under imperfect competition, 194–9
Marginal revenue, 167–172, 194–200
Marginal utility, defined, 127–128; measurement of, 128–32; curve of, 131–32; law of, 132–33; of durable goods, 133; relation to demand, 133–35; balancing of, by consumers, 144–45; dependence upon scarcity, 146; of money and progressive taxation, 487–90
Margin of cultivation, intensive and extensive, on urban lands, 37; defined, 400; extensive, 400; intensive, 400
Market, defined, 90–91; free, 91; clearing the, 111; retail egg, 119–22; classification of types of, 224–25; money, 414; foreign, 778–79; quotas, 733–34, 737
Markets, classification of, 192; prices of same good in different, 242–43
Matured economy, explanation of business cycle by theory of, 629–31
Marx, Karl, 801–806
Mediation, 708–10
Mercantilism, 75–76, 572–73; economic nationalism an aim of, 596; influence on tariff treaties, 603–604; and favorable balance of trade, 598
Mergers, 770, 773
Miller-Tydings Act of 1937, 781–82, 786–87
Minimum wage laws, state, 672; constitutionality of, 672; appraisal of, 673–74
Mining, increasing costs in, 177
Monetary problems, contemporary international, 575–94
Money, defined, 15, 253; primitive,

252; characteristics of, 252–53; principles of, 252–78; a medium of exchange, 253–54; a standard of value, 254; a store of value, 254; government control of, 254–59; standard, 256–57; irredeemable, 257; classification of, 257–58; representative, 257, 260; credit, 258; token, 258; ideal qualities of, 259–60; gold and the value of, 260–61; bimetallic, 263–67; history of American, 270–78; American stock of, 275, 277–78; American, in circulation, 277–78; "lawful," 316; value of, 353–54; and prices, 353–73; variations in value of, 354; effect of trade on value of, 361; effect of quantity on value of, 361–63; velocity of circulation of, 363; forces determining value of, 363–66; quantity theory of, 363–66; effects of fluctuating value of, 367; managed, 368–70; and control of business cycle, 368–73; stable, 368–73; one hundred per cent, 371–72; rate of interest and quantity of, 414–16; in communist society, 814; in Russia, 825
Money economy, 88
Money system, standard, 257–60; irredeemable, 257; of Great Britain, 580–82
Monopoly, 192, 226, 769; and combination, 72; and competition, 226–27; and imperfect competition, 227–28; and the public interest, 228; curtailment of, 638; government regulation of, 770–79
Monopolistic industry. *See* Monopoly
Moratoria, of international indebtedness, 585–86; on farm mortgage debt, 726
Motor carriers, 748–49, 768
Mutual savings banks, number and importance, 351

National bank notes, 315–16
National banks, 315–18; notes of, 315–16; members of federal reserve, 319–21; and branch banking, 333–34, 338–39; capital requirements of, 339–40; loans of, on real estate, 349; statistics of, 351
National Credit Corporation, 341–42
National Housing act, 345; real estate loans under, 349
National Industrial Conference Board, on wages of women, 670
National Industrial Recovery Act, 654, 780–83; and child labor, 665; minimum wage provisions of codes of, 672; and company unions, 691; settlement of labor disputes under, 708; objectives of, 782–83; and wages, 782–83; effects of, 782–83; unconstitutionality of, 782–83; labor provisions of 782–83
Nationalism, economic, and the protective tariff, 595–613; nature of economic, 596; economic, during Second World War, 612; in Russia, 819
National Labor Relations Act, 688–89, 708; amended by Labor Management Relations Act, 690; effect on company unions, 691; collective bargaining provisions of, 697–98
See also Labor Management Relations Act
National Mediation Board, 709–10
Net profit, 53; from operations, 52
Newlands act, 709
New York Stock Exchange, 96–97, 174; price determination on, 102–12
New Zealand, compulsory arbitration in, 714–18
Notes, promissory, 284–95; proceeds of, 286; endorsed, 287; secured or unsecured, 287; bank, 306–308; of banks, compared with deposit, 306–307; federal reserve, 308, 322–24
See also Bank notes; Discount

Oligopoly, 191–92
One hundred per cent money, 371–72

Parity, and income, 725, 741–42; price, 728–29; payments, 733
Partnership, 36–38; unlimited liability of, 37, 37n
Patents, 789
Payrolls, index of factory, 618; tax on, 658–59
Pensions, for aged, 674–75
Picketing, 705–706

Planned economy, in Russia, 819–21; judgment of, 831–32
Planning, under NIRA, 783
Pool, a monopoly device, 772
Population, influence of taxation on location of, 500
Pound, sterling, stabilization of, 580–82
Power, and large-scale production, 63
Preferred stock, 42
Present worth, 233–34
Price, defined, 15; distinguished from value, 15–16; not determined by profits, 50; and control of production, 84–85; consumers' influence on, 85, 90; determined by bargaining, 87; and the market, 88–99; "just," 89; legal determination of, 89; customary, 89–90; quoted by seller, 98; quoted by buyer, 99; and demand, 102–103, 118–19; determination of, under pure competition, 102–12; and supply, 106–107, 118–19, 175–80; equilibrium, 111–12; basic law of change in, 122; marginal utility and, 145–46; anomalies resolved, 149–46; and demand under imperfect competition, 192–94; under imperfect competition and monopoly, 211–28; determination under imperfect competition, 212–23; and output under imperfect competition, 217–19; cutting of, 218–19; agreements, 219, 221; "leadership," 219; control of monopolist over, 226–28; discount and laws of, 238–40, of same good in different markets, 242–43; rates of interest, wages, rent and, 377–78; in agriculture, 396; relation of rent to, 404–405; parity, 728–29, 733, 741–42; "floors" and "ceilings," 738–40; prohibition of discrimination in, 777, 786
Price control, 727; safeguards against price cutting under imperfect competition, 219; economic obstacles to, 219–21
Price level, 355–73, 727; measured by index numbers, 355–59; forces determining, 363; quantity of money and, 364–70; income and fluctuations of, 367; and business cycle, 368–73; and demand for loans, 415–16; and exchange, 566, 566n; effect of gold movement on, 566; and gold imports, 572; measurement for purchasing power par of exchange, 579; stability of domestic, 583–84; and purchasing power par of exchange, 777
Price relatives, 355–59
Prices, interrelationship of, 231–50; of competing goods, 243–44; of complementary goods, 244, of competing cost goods, 244–45; of joint cost goods, 245; of "tandem goods," 245–46; and money, 353–73; fluctuations of, 353–73; international, 539–40; and stable economy, 638; in agriculture, 721–23, 725–28, 731, 738–42
See also Competitive price; Price
Private property, 78; socialist criticism of, 797–800
Processing taxes, 729–30
Production, 18–33; nature of, 18–19; defined, 19–20; factors of, 20; direct, 21–22; indirect, or roundabout, 22; factors of, no corresponding class division, 27; the standard of living and large-scale, 56–57; and growth of capital, 57–59; place of small scale, 64–65; control of, 75, 84–87, motive for, 83–84; to order, 155; for a general market, 155; effect of costs on, 171–72; price cutting and marginal, 220–21; distribution in the process of, 376–77; business demand for factors of, 378–81; control of agricultural, 729, 731, 740; of agricultural products during Second World War, 740; socialist criticism of, 798–810, 803–804; evaluation of present system of, 831–32
See also Business cycle
Production, large-scale. *See* Large-scale production
Productive factors, demand for, 371–92
Profit and loss statement, 51–54, 441–42
Profit, net, 51–52
Profits, 441–51; concealing through stock-watering, 50–51; under imperfect competition, 197–99; defined, 376, 443; and risk, 392; nature of,

392; a residual income, 443–44; promoter's, 445–46; accrue to owners, 446; and losses of corporations, 446–49; and risk, 445–50; effect of competition on, 449–50; criticism of, 450–51; savings of, 790–91
Progressive taxation, 492–96; use to effect proportional taxation, 497–98; and marginal utility of money, 486–92
See also Taxation
Promotion, of new industries, 445–6
Property, defined, 11; division of, 11–12; in free persons, 12–13; distinguished from wealth, 13, evidence of, 13–14; and capital, 232–33; valuation of, 249–50, 463–66; taxable, 463–64; under-valuation and its effect on taxation, 465–66; confusion with wealth in property tax, 466
See also Private property
Property account, explained, 45–49; surplus in, 46–47
Property tax. *See* General property tax
Proportional taxation, 497–98
Proportionality, law of, 438–40
Prosperity, in business cycle, 622–24; limiting factors in, 623; to save and spend, 629–30
Protection, defined, 474; in Great Britain, 476, 603; in United States, 476–77; non-fiscal taxation, 505; pros and cons of, 596–611
See also Customs tariff; Protective tariff
Protective tariff, 596–611; economic effects of, 473–75; fiscal weakness of, 476–77; and economic nationalism, 595–613; gold argument for, 598; favorable balance of trade argument for, 598; added industry argument, 599–600; and wages, 600–601; wages and labor cost, 601; alleged danger of withdrawing, 601–602; alleged saving in transportation cost, 602–603; alleged need to meet tariffs of other nations, 603–604; alleged gain from injuring other nations, 604; territorial unit for, 604–605; benefits of foreign competition, 605; necessity of gradual reduction of, 606; temporary protection, 606; infant industry protection, 606–607; use against dumping, 607–608; military reasons for, 609–610; bounties compared with, 610–11; ship subsidies compared with, 611; preferential arrangements in, 611; British preferential arrangements in, 611; U. S. reciprocal trade agreements, 611–12; International Conference on Trade and Employment, 611–12; and stable economy, 639
Public finance, American, 508–30
Public utilities, legal determinations of rates of, 89; valuation of, 249–50
Public Utility Holding Company Act of 1935, 771–72
Public works, and unemployment, 653–55; program of the United States, 654–55
"Pump-priming," 632–33
Purchasing power par, control of exchange rates with respect to, 579
Pure competition, 100–101, 186, 192
See also Competition; Free economy

Railroads, early development of, 746–48; growth of, 746–48; topographical limitations of, 746–48, costs of, 749–52; and labor costs, 750; government regulation of, 753–54; geography and rate competition of, 756–57
Railroad rates, 754–67: bargaining of shippers, 758; determination of specific, 758; "class," 759
Railway Labor Act, 709–10
Rationing, by price, 86
Real estate, loans by banks on, 349; taxation of, 467–68
Recession, in business cycle, 617, 624–25
Reciprocal trade agreements of the U. S., 611–12
Reconstruction Finance Corporation, 341–43
Recovery, under AAA, 731–32; under NIRA, 783
Rediscount, by federal reserve banks, 321–22
Reform, status of monetary, 372; to eliminate rent, 407–408; of profit system, 450

Relief, of unemployment by labor unions 655–56; state and local, 660; Ghent system of unemployment, 656
Rent, economic, 374–75, 393–468; rates of, 377–78; explicit and implicit, 392, 397–98, 442–43; in agriculture, 394–401; relation to prices, 396, 404–405; and producer's surplus, 396–97; of urban lands, 401–403; site, 401; and land value, 403–404; quasi-, 405–406; distinguished from other income, 406; taxation of, 407–408; shifting of tax on, 480–81
Representative money, 257, 260, 262, 267
Research, and large-scale production, 63
Reserve, bank, 300–305; function of bank, 300–301
Reserve ratios, control of bank, 301–306
Reserves, problem of national bank, 317–18; centralized bank, 318; of federal reserve banks, 326–27; centralized, 326–27; mobile, 327; of banks and rate of interest, 415–16
Restrictive practices, of producers, 637; of labor, 636, 700–701
Revenue. *See* Marginal revenue
Revenues, of U. S. government, 527–30 *See also* Government revenues
Revival, in business cycle, 616; causes of business, 621–22; and prosperity, 621–22
Risk, 26; of security holders, 44; and interest, 417–18; profit and, 441–51; nature of business, 444; reducing of, 444
Robinson-Patman Act of 1936, 786
Russia, government regulation in, 75

Sales problem, immediate, 156, 175; remote, 156, 175
Sales promotion, 205–207
Sales tax, 473
Saving, 23–25; and interest rate, 413–14; over-, 423; source of capital formation, 410–11; incentives for, 411–12
Savings, by individuals, 411–12; corporate, 412–13; alleged increase in, 629–31; in Russia, 824, 830–31
Seasonal variation, 615–16
Secular trend, 615
Securities act of 1933, 771
Securities act of 1934, 771
Securities, regulation of issue of, 771
Seigniorage, defined, 275n
Seller's policy, under imperfect competition, 186–208
Sellers, and their costs, 154–66; agreements by, 219
Sherman Anti-trust Act of 1890, 762, 770, 774–78
Shifting and incidence, of taxation, 479–81
Short run, seller's policy in, 171; sales problem, 175; supply in, 180–81; equilibrium under imperfect competition, 212–13
Silver, why used as money, 259–60; ratio of, to gold under bimetallism, 263–64; as standard money, 263–67; domestic purchases of, 274–76; federal government's stock of, 275, 277–78; London conference's agreement concerning, 274; Thomas amendment and, 274; nationalization of, 275; ratio of, to gold held by U. S. Treasury, 275; act of June, 1934, concerning purchase of, 275
Silver certificates, United States, 275
Single tax, 408
"Sit-down" strike, 706
Socialism, 796–832; hostility to free enterprise, 81; programs of, 797, 806–807; theoretical bases of, 797–98; defined, 798; guild, 806; state, 806; tactics, 807–809; political action of, 808–809; appraisal of, 809–12; capital formation under, 812
See also Communism; Fascism
Social Security Act, 649–50, 658–60, 675–77
Soil conservation, in agriculture, 732–33, 736–37
Soil Conservation and Domestic Allotment Act of 1936, 732
Soil Conservation Service, 736
Specialization, and small-scale production, 65; within combination, 69; weaknesses of, within combination, 72–74; causes of regional, 532; gain from, 532–33

Spending, during 1930's, 629–37; government deficit, 630
Spreading the work, 642
Stabilization fund, for foreign exchange, 369–70, 589
Stabilized dollar, 370
Standard of living, rise of the, 56–57; relation to wage rates, 432
Standardization, a prerequisite of large-scale production, 61
Standard money, 257
See also Money
State banks, number and importance, 351
See also Banks
Steagall amendment, 739
"Sterling area," 588–89
Stock, common, 42; dividends on, 42; preferred, 42–43; cumulative preferred, 42; noncumulative preferred, 42; participating preferred, 42–43; and bonds compared, 44; par value, 48; book value, 48; market value, 48–49; watered, 50–51
See also Capital stock
Stock dividends, 49–50
Strikes, "sit-down," 686, 706; of labor unions, 701–706, statistics of U. S., 702–703; sympathetic and general, 703–704; legal status of, 705–706; devices to prevent or settle, 706–718
Subsidies, ship, 11
Substitutes, relation to competition and monopoly, 227
Supply, 167–84; control by rationing, 86–87; and price, 106–107; schedule of, 107–108; and demand together, 107–11; effect of decreased, 111–12, 118–19; effect of increased, 111–12, 118–19; laws of demand and, 111–12; two possible meanings, 114–15; effect of changes on price and quantity exchanged on demand and, 115–19; governed by cost, 156–57; relation of marginal cost curve to curve of, 168–70; under competition, 168–84; individual, 172–73; defined, 173; and stock, 174–75; law of, 180; long-run and short-run, 180; elasticity of, 181–82; discount partly determines, 238–40; of loan funds, 411–14; of labor, 426–27
Supreme Court, and child labor legislation, 667; and National Labor Relations acts, 689; and compulsory arbitration, 715–16
Surplus, in property account, 46–47; producer's, 396–97, 401; government, 455; commodity, 736
Syndicalists, 806

"Tandem goods," prices of, 245–46
Tangible personal property, taxation of, 467
Tariff, protective, 473–77; prohibitive, 474–75; revenue, 476–77
See also Protective tariff
Tax, defined, 456; distinguished from fees, 456–57; object or base of, 458; immediate payer of, 458; subject of, 458; source of, 458
Taxation, valuation in, 249; characteristics of, 456; shifting and incidence of, 458; direct, 458–59; indirect, 458–59; technical definitions, 458–59; proportional, 459, 497–98; progressive, 459–61, 492–96; regressive, 460, 497–98; degressive, 460; principle forms of, 461; escape from, 465–66; double, 466, 492; of corporations, 466–67, 494; principles of, 482–507; an attribute of sovereignty, 482–83; powers of, 482–84; purpose of, 484; *quid pro quo* of, 484–85; distribution of burden of, 485–86; benefit principle of, 486; cost principle of, 486; ability to pay principle of, 486–91; theories of sacrifice of, 487–91; and rule of equity, 492; ability to pay and progressive, 493; revenue aspects of progressive, 495–96; a neutral system of, 496–99; ability to pay and proportional, 498; tax legislation and theories of, 499; collateral effects of, 499–507; and geographical shifting of industry, 500; effect of form of, 500; effect on business expenditures, 500–501; effect on location of population, 500; effect on charitable contributions, 501; "incentive," 502; effect on industry, 502; effect on business opportunity, 502–503; effect of heavy, 503; cost of compliance with laws of, 504; non-

fiscal, 504–505; and Constitution of United States, 506–507; federal excise, 515–17; of payrolls, 658–59
See also Income tax; Tariff; Taxes
Taxes, apportionment of, 462–63; income, 469–72; consumption, 472–77; excise, 472–73, 515–17; sales, 473; inheritance, 477–78; estate, 477–78; death, 477–78; business, 478–79; license, 478; poll, 479; capitalization of, 481; progressive, 492–96; sumptuary, 504–505; processing, 729–30, 732
Tax rates, relation of, to base, 459–60; progressive, 459–61; determination of property, 462–63; and income, 469, 471
Technological change, and unemployment, 643–45
Time preference, 383–86, 410–11; and insurance, 384, 384n; rate of, 385; and borrowing for consumption, 385–86; and capital formation, 410–11
Trade, gain resulting from, 136–37; effect of, on value of money, 361; interregional, 531–54; bases of interregional, 531–32; multilateral, 537–38, 552; bilateral, 538–39, 548; and international prices, 539–40; world, 541–42; United States foreign, 542–48; government control of, 548–49; "favorable" and "unfavorable" balances of, 570–73; government action to correct unfavorable balances of, 585–87; restrictions on international, 595–97, 771, 777–79; use of protective tariff with state controlled, 609; restriction during Second World War of foreign, 612; foreign export, 778–79
See also Exports; Imports; Free Trade
Trade Agreements act of 1934, 611
Trade Union. *See* Labor union
Transfer, defined, 14
Transportation, 744–68; economics of, 744; and topographic factors, 744–49; canal system of, 745–46; railroad system of, 746–48; highway system of, 748–49; cost of, 749–52; and competition between carriers, 752–58; government regulation of, 753–54, 762; and rate associations, 760–62; place of the various forms of, 767–68
Transportation act of 1920, 708, 764–66
Transportation act of 1933, 765–67
Transportation rates, competition in, 755–58; "commodity," 759–60; discrimination in, 760; process of setting, 760–62; government regulation of, 762–66; basic standards for, 763–64; general level of, 764–67
Tripartite agreement, 589–90
Trusts, legislation against, in U. S., 770–71, 772–79; history of, in U. S., 772–73; defined, 772–73n; limitation of laws against, 775–77
Tying contract, 775

"Unearned increment," 407
See also Rent
Unemployment, 640–60; meaning of, 640; extent of, 640–42; causes of, 642–46; technological, 643–45; seasonal, 645–46; cyclical, 646; remedies for, 648–55; insurance against, 655–60
See also Employment
Unemployment insurance, in Great Britain, 656–58; in U. S., 658–60; state laws, 658–60
Unfair labor practices, 688, 690; and company unions, 691
Union. *See* Company unions; Labor unions
United Mine Workers, 685, 694
Union shop, 700
United States Steel Corporation, 71–73, 686
Usury, laws of, 422
Utility, defined, 6–7, 127; creation of, 18–19; forms of, 19; subjectively viewed, 126–27
See also Marginal utility

Variable cost. *See* Cost
Valuation, in accounting, 247–48; public utility, 248–49; tax, 249–50; of property, 463–66
Value, the common unit of measure, 14; defined, 15, 16, 353n; distin-

guished from price, 15–16; book, 48; par, 48; market, 48–49; practical aspects of theory of, 246–50; money a standard of, 353–55; of money, 353–55; of agricultural land, 737–38

Wages, 425–40; defined, 4, 375; rate of, 7, 377–78; implicit, 392; differences among groups of workers, 425; forces determining, 425–29; law of, 427–28, 673–74; based on discounted marginal productivity, 427–29; sphere of bargaining in, 429–32; and bargaining power of workers, 430–32; individual variations in, 432–33; geographical variations in, 433–34; of self-employed workers, 435; of non-business workers, 435–36; and the control of social classes, 437–38; and protection, 600–601; and flexibility of employment, 651; lag in adjustment of, 651; of women and children, 670–71; under Fair Labor Standards Act, 672–73; legal minimum, 672–73; in Russia 823

Wages of management, 442–43; and interest, 418

Wagner-Peyser act, 649–50

Wants, 1–3; without limit, 3–4; limited means for satisfying, 4

War Labor Board, 716–17

Watson-Parker act, 708

Wealth, defined, 7–9; measurement of, 10; distinguished from income, 10; distinguished from property, 13; production of 19; confused with property in taxation, 380–81; gold and U. S., 599

see also Distribution

Webb-Pomerene act of 1918, 778–79

Wheat, production and prices, 721–22, 728–32

Wheeler-Lea act of 1938, 778, 780

Woman labor, 667–71; extent and character of, 668–69; bargaining power of, 670–71; wages of, 670–71